CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS INVOLVED IN HOMEOSTASIS AND ADAPTATION

Related Applications

This application claims priority to prior filed U.S. Provisional Patent Application Serial No. 60/141031, filed June 25, 1999, U.S. Provisional Patent Application Serial No. 60/143694, filed July 14, 2000, and U.S. Provisional Patent Application Serial No. 60/151778, filed August 31, 1999. This application also claims priority to German Application No. 19931418.7, filed July 8, 1999, German Application No. 19932124.8, filed July 9, 1999, German Application No. 19932126.4, filed July 9, 1999, German Application No. 19932127.2, filed July 9, 1999, German Application No. 19932133.7, filed July 9, 1999, German Application No. 19932207.4, filed July 9, 1999, German Application No. 19932208.2, filed July 9, 1999, German Application No. 19932225.2, filed July 9, 1999, German Application No. 19932229.5, filed July 9, 1999, German Application No. 19932914.1, filed July 9, 1999, German Application No. 19933006.9, filed July 9, 1999, German Application No. 19940765.7, filed August 27, 1999, German Application No. 19940768.1, filed August 27, 1999, German Application No. 19940831.9, filed August 27, 1999, German Application No. 19940832.7, filed August 27, 1999, German Application No. 19941385.1, filed August 31, 1999, German Application No. 19941396.7, filed August 31, 1999, and German Application No. 19942087.4, filed September 3, 1999. The entire contents of all of the aforementioned applications are hereby expressly incorporated herein by this reference.

Background of the Invention

Certain products and by-products of naturally-occurring metabolic processes in cells have utility in a wide array of industries, including the food, feed, cosmetics, and pharmaceutical industries. These molecules, collectively termed 'fine chemicals', include organic acids, both proteinogenic and non-proteinogenic amino acids, nucleotides and nucleosides, lipids and fatty acids, diols, carbohydrates, aromatic compounds, vitamins and cofactors, and enzymes. Their production is most conveniently performed through the large-scale culture of bacteria developed to produce and secrete large quantities of one or more desired molecules. One particularly useful organism for this purpose is *Corynebacterium glutamicum*, a gram positive, nonpathogenic bacterium. Through strain selection, a number of mutant strains have been developed which produce an array of desirable compounds. However, selection of strains improved for the production of a particular molecule is a time-consuming and difficult process.

Summary of the Invention

The invention provides novel bacterial nucleic acid molecules which have a variety of uses. These uses include the identification of microorganisms which can be used to produce fine chemicals, the modulation of fine chemical production in *C. glutamicum* or related bacteria, the typing or identification of *C. glutamicum* or related bacteria, as reference points for mapping the *C. glutamicum* genome, and as markers for transformation. These novel nucleic acid molecules encode proteins, referred to herein as homeostasis and adaptation (HA) proteins.

C. glutamicum is a gram positive, aerobic bacterium which is commonly used in industry for the large-scale production of a variety of fine chemicals, and also for the degradation of hydrocarbons (such as in petroleum spills) and for the oxidation of terpenoids. The HA nucleic acid molecules of the invention, therefore, can be used to identify microorganisms which can be used to produce fine chemicals, *e.g.*, by fermentation processes. Modulation of the expression of the HA nucleic acids of the invention, or modification of the sequence of the HA nucleic acid molecules of the invention, can be used to modulate the production of one or more fine chemicals from a microorganism (*e.g.*, to improve the yield or production of one or more fine chemicals from a Corynebacterium or Brevibacterium species).

The HA nucleic acids of the invention may also be used to identify an organism as being *Corynebacterium glutamicum* or a close relative thereof, or to identify the presence of *C. glutamicum* or a relative thereof in a mixed population of microorganisms. The invention provides the nucleic acid sequences of a number of *C. glutamicum* genes; by probing the extracted genomic DNA of a culture of a unique or mixed population of microorganisms under stringent conditions with a probe spanning a region of a *C. glutamicum* gene which is unique to this organism, one can ascertain whether this organism is present. Although *Corynebacterium glutamicum* itself is nonpathogenic, it is related to species pathogenic in humans, such as *Corynebacterium diphtheriae* (the causative agent of diphtheria); the detection of such organisms is of significant clinical relevance.

The HA nucleic acid molecules of the invention may also serve as reference points for mapping of the *C. glutamicum* genome, or of genomes of related organisms. Similarly, these molecules, or variants or portions thereof, may serve as markers for genetically engineered Corynebacterium or Brevibacterium species.

The HA proteins encoded by the novel nucleic acid molecules of the invention are capable of, for example, performing a function involved in the maintenance of homeostasis in *C. glutamicum*, or in the ability of this microorganism to adapt to

different environmental conditions. Given the availability of cloning vectors for use in *Corynebacterium glutamicum*, such as those disclosed in Sinskey *et al.*, U.S. Patent No. 4,649,119, and techniques for genetic manipulation of *C. glutamicum* and the related *Brevibacterium* species (*e.g.*, *lactofermentum*) (Yoshihama et al, *J. Bacteriol.* 162: 591-597 (1985); Katsumata *et al.*, *J. Bacteriol.* 159: 306-311 (1984); and Santamaria *et al.*, *J. Gen. Microbiol.* 130: 2237-2246 (1984)), the nucleic acid molecules of the invention may be utilized in the genetic engineering of this organism to make it a better or more efficient producer of one or more fine chemicals. This improved production or efficiency of production of a fine chemical may be due to a direct effect of manipulation of a gene of the invention, or it may be due to an indirect effect of such manipulation.

There are a number of mechanisms by which the alteration of an HA protein of the invention may directly affect the yield, production, and/or efficiency of production of a fine chemical from a *C. glutamicum* strain incorporating such an altered protein. For example, by engineering enzymes which modify or degrade aromatic or aliphatic compounds such that these enzymes are increased or decreased in activity or number, it may be possible to modulate the production of one or more fine chemicals which are the modification or degradation products of these compounds. Similarly, enzymes involved in the metabolism of inorganic compounds provide key molecules (*e.g.* phosphorous, sulfur, and nitrogen molecules) for the biosynthesis of such fine chemicals as amino acids, vitamins, and nucleic acids. By altering the activity or number of these enzymes in *C. glutamicum*, it may be possible to increase the conversion of these inorganic compounds (or to use alternate inorganic compounds) to thus permit improved rates of incorporation of inorganic atoms into these fine chemicals. Genetic engineering of *C. glutamicum* enzymes involved in general cellular processes may also directly improve fine chemical production, since many of these enzymes directly modify fine chemicals (*e.g.*, amino acids) or the enzymes which are involved in fine chemical synthesis or secretion. Modulation of the activity or number of cellular proteases may also have a direct effect on fine chemical production, since many proteases may degrade fine chemicals or enzymes involved in fine chemical production or breakdown.

Further, the aforementioned enzymes which participate in aromatic/aliphatic compound modification or degradation, general biocatalysis, inorganic compound metabolism or proteolysis are each themselves fine chemicals, desirable for their activity in various *in vitro* industrial applications. By altering the number of copies of the gene for one or more of these enzymes in *C. glutamicum* it may be possible to increase the number of these proteins produced by the cell, thereby increasing the potential yield or efficiency of production of these proteins from large-scale *C. glutamicum* or related bacterial cultures.

The alteration of an HA protein of the invention may also indirectly affect the yield, production, and/or efficiency of production of a fine chemical from a *C. glutamicum* strain incorporating such an altered protein. For example, by modulating the activity and/or number of those proteins involved in the construction or rearrangement of the cell wall, it may be possible to modify the structure of the cell wall itself such that the cell is able to better withstand the mechanical and other stresses present during large-scale fermentative culture. Also, large-scale growth of *C. glutamicum* requires significant cell wall production. Modulation of the activity or number of cell wall biosynthetic or degradative enzymes may allow more rapid rates of cell wall biosynthesis, which in turn may permit increased growth rates of this microorganism in culture and thereby increase the number of cells producing the desired fine chemical.

By modifying the HA enzymes of the invention, one may also indirectly impact the yield, production, or efficiency of production of one or more fine chemicals from *C. glutamicum*. For example, many of the general enzymes in *C. glutamicum* may have a significant impact on global cellular processes (*e.g.*, regulatory processes) which in turn have a significant effect on fine chemical metabolism. Similarly, proteases, enzymes which modify or degrade possibly toxic aromatic or aliphatic compounds, and enzymes which promote the metabolism of inorganic compounds all serve to increase the viability of *C. glutamicum*. The proteases aid in the selective removal of misfolded or misregulated proteins, such as those that might occur under the relatively stressful environmental conditions encountered during large-scale fermentor culture. By altering these proteins, it may be possible to further enhance this activity and to improve the viability of *C. glutamicum* in culture. The aromatic/aliphatic modification or degradation proteins not only serve to detoxify these waste compounds (which may be encountered as impurities in culture medium or as waste products from cells themselves), but also to permit the cells to utilize alternate carbon sources if the optimal carbon source is limiting in the culture. By increasing their number and/or activity, the survival of *C. glutamicum* cells in culture may be enhanced. The inorganic metabolism proteins of the invention supply the cell with inorganic molecules required for all protein and nucleotide (among others) synthesis, and thus are critical for the overall viability of the cell. An increase in the number of viable cells producing one or more desired fine chemicals in large-scale culture should result in a concomitant increase in the yield, production, and/or efficiency of production of the fine chemical in the culture.

The invention provides novel nucleic acid molecules which encode proteins, referred to herein as HA proteins, which are capable of, for example, performing a function involved in the maintenance of homeostasis in *C. glutamicum*, or of

participating in the ability of this microorganism to adapt to different environmental conditions. Nucleic acid molecules encoding an HA protein are referred to herein as HA nucleic acid molecules. In a preferred embodiment, an HA protein participates in *C. glutamicum* cell wall biosynthesis or rearrangements, metabolism of inorganic compounds, modification or degradation of aromatic or aliphatic compounds, or possesses a *C. glutamicum* enzymatic or proteolytic activity. Examples of such proteins include those encoded by the genes set forth in Table 1.

Accordingly, one aspect of the invention pertains to isolated nucleic acid molecules (*e.g.*, cDNAs, DNAs, or RNAs) comprising a nucleotide sequence encoding an HA protein or biologically active portions thereof, as well as nucleic acid fragments suitable as primers or hybridization probes for the detection or amplification of HA-encoding nucleic acids (*e.g.*, DNA or mRNA). In particularly preferred embodiments, the isolated nucleic acid molecule comprises one of the nucleotide sequences set forth in Appendix A or the coding region or a complement thereof of one of these nucleotide sequences. In other particularly preferred embodiments, the isolated nucleic acid molecule of the invention comprises a nucleotide sequence which hybridizes to or is at least about 50%, preferably at least about 60%, more preferably at least about 70%, 80% or 90%, and even more preferably at least about 95%, 96%, 97%, 98%, 99% or more homologous to a nucleotide sequence set forth in Appendix A, or a portion thereof. In other preferred embodiments, the isolated nucleic acid molecule encodes one of the amino acid sequences set forth in Appendix B. The preferred HA proteins of the present invention also preferably possess at least one of the HA activities described herein.

In another embodiment, the isolated nucleic acid molecule encodes a protein or portion thereof wherein the protein or portion thereof includes an amino acid sequence which is sufficiently homologous to an amino acid sequence of Appendix B, *e.g.*, sufficiently homologous to an amino acid sequence of Appendix B such that the protein or portion thereof maintains an HA activity. Preferably, the protein or portion thereof encoded by the nucleic acid molecule maintains the ability to participate in the maintenance of homeostasis in *C. glutamicum*, or to perform a function involved in the adaptation of this microorganism to different environmental conditions. In one embodiment, the protein encoded by the nucleic acid molecule is at least about 50%, preferably at least about 60%, and more preferably at least about 70%, 80%, or 90% and most preferably at least about 95%, 96%, 97%, 98%, or 99% or more homologous to an amino acid sequence of Appendix B (*e.g.*, an entire amino acid sequence selected from those sequences set forth in Appendix B). In another preferred embodiment, the protein is a full length *C. glutamicum* protein which is substantially homologous to an entire

amino acid sequence of Appendix B (encoded by an open reading frame shown in Appendix A).

In another preferred embodiment, the isolated nucleic acid molecule is derived from *C. glutamicum* and encodes a protein (*e.g.*, an HA fusion protein) which includes a biologically active domain which is at least about 50% or more homologous to one of the amino acid sequences of Appendix B and is able to participate in the repair or recombination of DNA, in the transposition of genetic material, in gene expression (*i.e.*, the processes of transcription or translation), in protein folding, or in protein secretion in *Corynebacterium glutamicum*, or has one or more of the activities set forth in Table 1, and which also includes heterologous nucleic acid sequences encoding a heterologous polypeptide or regulatory regions.

In another embodiment, the isolated nucleic acid molecule is at least 15 nucleotides in length and hybridizes under stringent conditions to a nucleic acid molecule comprising a nucleotide sequence of Appendix A. Preferably, the isolated nucleic acid molecule corresponds to a naturally-occurring nucleic acid molecule. More preferably, the isolated nucleic acid encodes a naturally-occurring *C. glutamicum* HA protein, or a biologically active portion thereof.

Another aspect of the invention pertains to vectors, *e.g.*, recombinant expression vectors, containing the nucleic acid molecules of the invention, and host cells into which such vectors have been introduced. In one embodiment, such a host cell is used to produce an HA protein by culturing the host cell in a suitable medium. The HA protein can be then isolated from the medium or the host cell.

Yet another aspect of the invention pertains to a genetically altered microorganism in which an HA gene has been introduced or altered. In one embodiment, the genome of the microorganism has been altered by introduction of a nucleic acid molecule of the invention encoding wild-type or mutated HA sequence as a transgene. In another embodiment, an endogenous HA gene within the genome of the microorganism has been altered, *e.g.*, functionally disrupted, by homologous recombination with an altered HA gene. In another embodiment, an endogenous or introduced HA gene in a microorganism has been altered by one or more point mutations, deletions, or inversions, but still encodes a functional HA protein. In still another embodiment, one or more of the regulatory regions (*e.g.*, a promoter, repressor, or inducer) of an HA gene in a microorganism has been altered (*e.g.*, by deletion, truncation, inversion, or point mutation) such that the expression of the HA gene is modulated. In a preferred embodiment, the microorganism belongs to the genus *Corynebacterium* or *Brevibacterium*, with *Corynebacterium glutamicum* being particularly preferred. In a preferred embodiment, the microorganism is also utilized for

the production of a desired compound, such as an amino acid, with lysine being particularly preferred.

In another aspect, the invention provides a method of identifying the presence or activity of *Cornyebacterium diphtheriae* in a subject. This method includes detection of one or more of the nucleic acid or amino acid sequences of the invention (*e.g.*, the sequences set forth in Appendix A or Appendix B) in a subject, thereby detecting the presence or activity of *Corynebacterium diphtheriae* in the subject.

Still another aspect of the invention pertains to an isolated HA protein or a portion, *e.g.*, a biologically active portion, thereof. In a preferred embodiment, the isolated HA protein or portion thereof can participate in the maintenance of homeostasis in *C. glutamicum*, or can perform a function involved in the adaptation of this microorganism to different environmental conditions. In another preferred embodiment, the isolated HA protein or portion thereof is sufficiently homologous to an amino acid sequence of Appendix B such that the protein or portion thereof maintains the ability to participate in the maintenance of homeostasis in *C. glutamicum*, or to perform a function involved in the adaptation of this microorganism to different environmental conditions.

The invention also provides an isolated preparation of an HA protein. In preferred embodiments, the HA protein comprises an amino acid sequence of Appendix B. In another preferred embodiment, the invention pertains to an isolated full length protein which is substantially homologous to an entire amino acid sequence of Appendix B (encoded by an open reading frame set forth in Appendix A). In yet another embodiment, the protein is at least about 50%, preferably at least about 60%, and more preferably at least about 70%, 80%, or 90%, and most preferably at least about 95%, 96%, 97%, 98%, or 99% or more homologous to an entire amino acid sequence of Appendix B. In other embodiments, the isolated HA protein comprises an amino acid sequence which is at least about 50% or more homologous to one of the amino acid sequences of Appendix B and is able to participate in the maintenance of homeostasis in *C. glutamicum*, or to perform a function involved in the adaptation of this microorganism to different environmental conditions, or has one or more of the activities set forth in Table 1.

Alternatively, the isolated HA protein can comprise an amino acid sequence which is encoded by a nucleotide sequence which hybridizes, *e.g.*, hybridizes under stringent conditions, or is at least about 50%, preferably at least about 60%, more preferably at least about 70%, 80%, or 90%, and even more preferably at least about 95%, 96%, 97%, 98,%, or 99% or more homologous, to a nucleotide sequence of Appendix B. It is also preferred that the preferred forms of HA proteins also have one or more of the HA bioactivities described herein.

The HA polypeptide, or a biologically active portion thereof, can be operatively linked to a non-HA polypeptide to form a fusion protein. In preferred embodiments, this fusion protein has an activity which differs from that of the HA protein alone. In other preferred embodiments, this fusion protein participates in the maintenance of homeostasis in *C. glutamicum*, or performs a function involved in the adaptation of this microorganism to different environmental conditions. In particularly preferred embodiments, integration of this fusion protein into a host cell modulates production of a desired compound from the cell.

In another aspect, the invention provides methods for screening molecules which modulate the activity of an HA protein, either by interacting with the protein itself or a substrate or binding partner of the HA protein, or by modulating the transcription or translation of an HA nucleic acid molecule of the invention.

Another aspect of the invention pertains to a method for producing a fine chemical. This method involves the culturing of a cell containing a vector directing the expression of an HA nucleic acid molecule of the invention, such that a fine chemical is produced. In a preferred embodiment, this method further includes the step of obtaining a cell containing such a vector, in which a cell is transfected with a vector directing the expression of an HA nucleic acid. In another preferred embodiment, this method further includes the step of recovering the fine chemical from the culture. In a particularly preferred embodiment, the cell is from the genus *Corynebacterium* or *Brevibacterium*, or is selected from those strains set forth in Table 3.

Another aspect of the invention pertains to methods for modulating production of a molecule from a microorganism. Such methods include contacting the cell with an agent which modulates HA protein activity or HA nucleic acid expression such that a cell associated activity is altered relative to this same activity in the absence of the agent. In a preferred embodiment, the cell is modulated for one or more *C. glutamicum* processes involved in cell wall biosynthesis or rearrangements, metabolism of inorganic compounds, modification or degradation of aromatic or aliphatic compounds, or enzymatic or proteolytic activities. The agent which modulates HA protein activity can be an agent which stimulates HA protein activity or HA nucleic acid expression. Examples of agents which stimulate HA protein activity or HA nucleic acid expression include small molecules, active HA proteins, and nucleic acids encoding HA proteins that have been introduced into the cell. Examples of agents which inhibit HA activity or expression include small molecules and antisense HA nucleic acid molecules.

Another aspect of the invention pertains to methods for modulating yields of a desired compound from a cell, involving the introduction of a wild-type or mutant HA gene into a cell, either maintained on a separate plasmid or integrated into the genome of

the host cell. If integrated into the genome, such integration can be random, or it can take place by homologous recombination such that the native gene is replaced by the introduced copy, causing the production of the desired compound from the cell to be modulated. In a preferred embodiment, said yields are increased. In another preferred embodiment, said chemical is a fine chemical. In a particularly preferred embodiment, said fine chemical is an amino acid. In especially preferred embodiments, said amino acid is L-lysine.

Detailed Description of the Invention

The present invention provides HA nucleic acid and protein molecules which are involved in *C. glutamicum* cell wall biosynthesis or rearrangements, metabolism of inorganic compounds, modification or degradation of aromatic or aliphatic compounds, or that have a *C. glutamicum* enzymatic or proteolytic activity. The molecules of the invention may be utilized in the modulation of production of fine chemicals from microorganisms, such as *C. glutamicum*, either directly (*e.g.*, where overexpression or optimization of activity of a protein involved in the production of a fine chemical (*e.g.*, an enzyme) has a direct impact on the yield, production, and/or efficiency of production of a fine chemical from the modified *C. glutamicum*), or an indirect impact which nonetheless results in an increase of yield, production, and/or efficiency of production of the desired compound (*e.g.*, where modulation of the activity or number of copies of a *C. glutamicum* aromatic or aliphatic modification or degradation protein results in an increase in the viability of *C. glutamicum* cells, which in turn permits increased production in a large-scale culture setting). Aspects of the invention are further explicated below.

I. Fine Chemicals

The term 'fine chemical' is art-recognized and includes molecules produced by an organism which have applications in various industries, such as, but not limited to, the pharmaceutical, agriculture, and cosmetics industries. Such compounds include organic acids, such as tartaric acid, itaconic acid, and diaminopimelic acid, both proteinogenic and non-proteinogenic amino acids, purine and pyrimidine bases, nucleosides, and nucleotides (as described *e.g.* in Kuninaka, A. (1996) Nucleotides and related compounds, p. 561-612, in Biotechnology vol. 6, Rehm *et al.*, eds. VCH: Weinheim, and references contained therein), lipids, both saturated and unsaturated fatty acids (*e.g.*, arachidonic acid), diols (*e.g.*, propane diol, and butane diol), carbohydrates (*e.g.*, hyaluronic acid and trehalose), aromatic compounds (*e.g.*, aromatic amines, vanillin, and indigo), vitamins and cofactors (as described in Ullmann's Encyclopedia of

Industrial Chemistry, vol. A27, "Vitamins", p. 443-613 (1996) VCH: Weinheim and references therein; and Ong, A.S., Niki, E. & Packer, L. (1995) "Nutrition, Lipids, Health, and Disease" Proceedings of the UNESCO/Confederation of Scientific and Technological Associations in Malaysia, and the Society for Free Radical Research – Asia, held Sept. 1-3, 1994 at Penang, Malaysia, AOCS Press, (1995)), enzymes, polyketides (Cane *et al.* (1998) *Science* 282: 63-68), and all other chemicals described in Gutcho (1983) Chemicals by Fermentation, Noyes Data Corporation, ISBN: 0818805086 and references therein. The metabolism and uses of certain of these fine chemicals are further explicated below.

A. *Amino Acid Metabolism and Uses*

Amino acids comprise the basic structural units of all proteins, and as such are essential for normal cellular functioning in all organisms. The term "amino acid" is art-recognized. The proteinogenic amino acids, of which there are 20 species, serve as structural units for proteins, in which they are linked by peptide bonds, while the nonproteinogenic amino acids (hundreds of which are known) are not normally found in proteins (see Ulmann's Encyclopedia of Industrial Chemistry, vol. A2, p. 57-97 VCH: Weinheim (1985)). Amino acids may be in the D- or L- optical configuration, though L-amino acids are generally the only type found in naturally-occurring proteins. Biosynthetic and degradative pathways of each of the 20 proteinogenic amino acids have been well characterized in both prokaryotic and eukaryotic cells (see, for example, Stryer, L. Biochemistry, 3rd edition, pages 578-590 (1988)). The 'essential' amino acids (histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan, and valine), so named because they are generally a nutritional requirement due to the complexity of their biosyntheses, are readily converted by simple biosynthetic pathways to the remaining 11 'nonessential' amino acids (alanine, arginine, asparagine, aspartate, cysteine, glutamate, glutamine, glycine, proline, serine, and tyrosine). Higher animals do retain the ability to synthesize some of these amino acids, but the essential amino acids must be supplied from the diet in order for normal protein synthesis to occur.

Aside from their function in protein biosynthesis, these amino acids are interesting chemicals in their own right, and many have been found to have various applications in the food, feed, chemical, cosmetics, agriculture, and pharmaceutical industries. Lysine is an important amino acid in the nutrition not only of humans, but also of monogastric animals such as poultry and swine. Glutamate is most commonly used as a flavor additive (mono-sodium glutamate, MSG) and is widely used throughout the food industry, as are aspartate, phenylalanine, glycine, and cysteine. Glycine, L-methionine and tryptophan are all utilized in the pharmaceutical industry. Glutamine,

valine, leucine, isoleucine, histidine, arginine, proline, serine and alanine are of use in both the pharmaceutical and cosmetics industries. Threonine, tryptophan, and D/ L-methionine are common feed additives. (Leuchtenberger, W. (1996) Amino aids – technical production and use, p. 466-502 in Rehm *et al.* (eds.) Biotechnology vol. 6, chapter 14a, VCH: Weinheim). Additionally, these amino acids have been found to be useful as precursors for the synthesis of synthetic amino acids and proteins, such as N-acetylcysteine, S-carboxymethyl-L-cysteine, (S)-5-hydroxytryptophan, and others described in Ulmann's Encyclopedia of Industrial Chemistry, vol. A2, p. 57-97, VCH: Weinheim, 1985.

The biosynthesis of these natural amino acids in organisms capable of producing them, such as bacteria, has been well characterized (for review of bacterial amino acid biosynthesis and regulation thereof, see Umbarger, H.E.(1978) *Ann. Rev. Biochem.* 47: 533-606). Glutamate is synthesized by the reductive amination of α-ketoglutarate, an intermediate in the citric acid cycle. Glutamine, proline, and arginine are each subsequently produced from glutamate. The biosynthesis of serine is a three-step process beginning with 3-phosphoglycerate (an intermediate in glycolysis), and resulting in this amino acid after oxidation, transamination, and hydrolysis steps. Both cysteine and glycine are produced from serine; the former by the condensation of homocysteine with serine, and the latter by the transferal of the side-chain β-carbon atom to tetrahydrofolate, in a reaction catalyzed by serine transhydroxymethylase. Phenylalanine, and tyrosine are synthesized from the glycolytic and pentose phosphate pathway precursors erythrose 4-phosphate and phosphoenolpyruvate in a 9-step biosynthetic pathway that differ only at the final two steps after synthesis of prephenate. Tryptophan is also produced from these two initial molecules, but its synthesis is an 11-step pathway. Tyrosine may also be synthesized from phenylalanine, in a reaction catalyzed by phenylalanine hydroxylase. Alanine, valine, and leucine are all biosynthetic products of pyruvate, the final product of glycolysis. Aspartate is formed from oxaloacetate, an intermediate of the citric acid cycle. Asparagine, methionine, threonine, and lysine are each produced by the conversion of aspartate. Isoleucine is formed from threonine. A complex 9-step pathway results in the production of histidine from 5-phosphoribosyl-1-pyrophosphate, an activated sugar.

Amino acids in excess of the protein synthesis needs of the cell cannot be stored, and are instead degraded to provide intermediates for the major metabolic pathways of the cell (for review see Stryer, L. Biochemistry 3rd ed. Ch. 21 "Amino Acid Degradation and the Urea Cycle" p. 495-516 (1988)). Although the cell is able to convert unwanted amino acids into useful metabolic intermediates, amino acid production is costly in terms of energy, precursor molecules, and the enzymes necessary to synthesize them.

Thus it is not surprising that amino acid biosynthesis is regulated by feedback inhibition, in which the presence of a particular amino acid serves to slow or entirely stop its own production (for overview of feedback mechanisms in amino acid biosynthetic pathways, see Stryer, L. Biochemistry, 3rd ed. Ch. 24: "Biosynthesis of Amino Acids and Heme" p. 575-600 (1988)). Thus, the output of any particular amino acid is limited by the amount of that amino acid present in the cell.

B. Vitamin, Cofactor, and Nutraceutical Metabolism and Uses

Vitamins, cofactors, and nutraceuticals comprise another group of molecules which the higher animals have lost the ability to synthesize and so must ingest, although they are readily synthesized by other organisms such as bacteria. These molecules are either bioactive substances themselves, or are precursors of biologically active substances which may serve as electron carriers or intermediates in a variety of metabolic pathways. Aside from their nutritive value, these compounds also have significant industrial value as coloring agents, antioxidants, and catalysts or other processing aids. (For an overview of the structure, activity, and industrial applications of these compounds, see, for example, Ullman's Encyclopedia of Industrial Chemistry, "Vitamins" vol. A27, p. 443-613, VCH: Weinheim, 1996.) The term "vitamin" is art-recognized, and includes nutrients which are required by an organism for normal functioning, but which that organism cannot synthesize by itself. The group of vitamins may encompass cofactors and nutraceutical compounds. The language "cofactor" includes nonproteinaceous compounds required for a normal enzymatic activity to occur. Such compounds may be organic or inorganic; the cofactor molecules of the invention are preferably organic. The term "nutraceutical" includes dietary supplements having health benefits in plants and animals, particularly humans. Examples of such molecules are vitamins, antioxidants, and also certain lipids (*e.g.*, polyunsaturated fatty acids).

The biosynthesis of these molecules in organisms capable of producing them, such as bacteria, has been largely characterized (Ullman's Encyclopedia of Industrial Chemistry, "Vitamins" vol. A27, p. 443-613, VCH: Weinheim, 1996; Michal, G. (1999) Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology, John Wiley & Sons; Ong, A.S., Niki, E. & Packer, L. (1995) "Nutrition, Lipids, Health, and Disease" Proceedings of the UNESCO/Confederation of Scientific and Technological Associations in Malaysia, and the Society for Free Radical Research – Asia, held Sept. 1-3, 1994 at Penang, Malaysia, AOCS Press: Champaign, IL X, 374 S).

Thiamin (vitamin B_1) is produced by the chemical coupling of pyrimidine and thiazole moieties. Riboflavin (vitamin B_2) is synthesized from guanosine-5'-triphosphate

(GTP) and ribose-5'-phosphate. Riboflavin, in turn, is utilized for the synthesis of flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD). The family of compounds collectively termed 'vitamin B_6' (*e.g.*, pyridoxine, pyridoxamine, pyridoxa-5'-phosphate, and the commercially used pyridoxin hydrochloride) are all derivatives of the common structural unit, 5-hydroxy-6-methylpyridine. Pantothenate (pantothenic acid, (R)-(+)-N-(2,4-dihydroxy-3,3-dimethyl-1-oxobutyl)-β-alanine) can be produced either by chemical synthesis or by fermentation. The final steps in pantothenate biosynthesis consist of the ATP-driven condensation of β-alanine and pantoic acid. The enzymes responsible for the biosynthesis steps for the conversion to pantoic acid, to β-alanine and for the condensation to panthotenic acid are known. The metabolically active form of pantothenate is Coenzyme A, for which the biosynthesis proceeds in 5 enzymatic steps. Pantothenate, pyridoxal-5'-phosphate, cysteine and ATP are the precursors of Coenzyme A. These enzymes not only catalyze the formation of panthothante, but also the production of (R)-pantoic acid, (R)-pantolacton, (R)-panthenol (provitamin B_5), pantetheine (and its derivatives) and coenzyme A.

Biotin biosynthesis from the precursor molecule pimeloyl-CoA in microorganisms has been studied in detail and several of the genes involved have been identified. Many of the corresponding proteins have been found to also be involved in Fe-cluster synthesis and are members of the nifS class of proteins. Lipoic acid is derived from octanoic acid, and serves as a coenzyme in energy metabolism, where it becomes part of the pyruvate dehydrogenase complex and the α-ketoglutarate dehydrogenase complex. The folates are a group of substances which are all derivatives of folic acid, which is turn is derived from L-glutamic acid, p-amino-benzoic acid and 6-methylpterin. The biosynthesis of folic acid and its derivatives, starting from the metabolism intermediates guanosine-5'-triphosphate (GTP), L-glutamic acid and p-amino-benzoic acid has been studied in detail in certain microorganisms.

Corrinoids (such as the cobalamines and particularly vitamin B_{12}) and porphyrines belong to a group of chemicals characterized by a tetrapyrole ring system. The biosynthesis of vitamin B_{12} is sufficiently complex that it has not yet been completely characterized, but many of the enzymes and substrates involved are now known. Nicotinic acid (nicotinate), and nicotinamide are pyridine derivatives which are also termed 'niacin'. Niacin is the precursor of the important coenzymes NAD (nicotinamide adenine dinucleotide) and NADP (nicotinamide adenine dinucleotide phosphate) and their reduced forms.

The large-scale production of these compounds has largely relied on cell-free chemical syntheses, though some of these chemicals have also been produced by large-scale culture of microorganisms, such as riboflavin, Vitamin B_6, pantothenate, and

biotin. Only Vitamin B_{12} is produced solely by fermentation, due to the complexity of its synthesis. *In vitro* methodologies require significant inputs of materials and time, often at great cost.

C. Purine, Pyrimidine, Nucleoside and Nucleotide Metabolism and Uses

Purine and pyrimidine metabolism genes and their corresponding proteins are important targets for the therapy of tumor diseases and viral infections. The language "purine" or "pyrimidine" includes the nitrogenous bases which are constituents of nucleic acids, co-enzymes, and nucleotides. The term "nucleotide" includes the basic structural units of nucleic acid molecules, which are comprised of a nitrogenous base, a pentose sugar (in the case of RNA, the sugar is ribose; in the case of DNA, the sugar is D-deoxyribose), and phosphoric acid. The language "nucleoside" includes molecules which serve as precursors to nucleotides, but which are lacking the phosphoric acid moiety that nucleotides possess. By inhibiting the biosynthesis of these molecules, or their mobilization to form nucleic acid molecules, it is possible to inhibit RNA and DNA synthesis; by inhibiting this activity in a fashion targeted to cancerous cells, the ability of tumor cells to divide and replicate may be inhibited. Additionally, there are nucleotides which do not form nucleic acid molecules, but rather serve as energy stores (*i.e.*, AMP) or as coenzymes (*i.e.*, FAD and NAD).

Several publications have described the use of these chemicals for these medical indications, by influencing purine and/or pyrimidine metabolism (*e.g.* Christopherson, R.I. and Lyons, S.D. (1990) "Potent inhibitors of *de novo* pyrimidine and purine biosynthesis as chemotherapeutic agents." *Med. Res. Reviews* 10: 505-548). Studies of enzymes involved in purine and pyrimidine metabolism have been focused on the development of new drugs which can be used, for example, as immunosuppressants or anti-proliferants (Smith, J.L., (1995) "Enzymes in nucleotide synthesis." *Curr. Opin. Struct. Biol.* 5: 752-757; (1995) *Biochem Soc. Transact.* 23: 877-902). However, purine and pyrimidine bases, nucleosides and nucleotides have other utilities: as intermediates in the biosynthesis of several fine chemicals (*e.g.*, thiamine, S-adenosyl-methionine, folates, or riboflavin), as energy carriers for the cell (*e.g.*, ATP or GTP), and for chemicals themselves, commonly used as flavor enhancers (*e.g.*, IMP or GMP) or for several medicinal applications (see, for example, Kuninaka, A. (1996) Nucleotides and Related Compounds in Biotechnology vol. 6, Rehm *et al.*, eds. VCH: Weinheim, p. 561-612). Also, enzymes involved in purine, pyrimidine, nucleoside, or nucleotide metabolism are increasingly serving as targets against which chemicals for crop protection, including fungicides, herbicides and insecticides, are developed.

The metabolism of these compounds in bacteria has been characterized (for reviews see, for example, Zalkin, H. and Dixon, J.E. (1992) "*de novo* purine nucleotide biosynthesis", in: Progress in Nucleic Acid Research and Molecular Biology, vol. 42, Academic Press:, p. 259-287; and Michal, G. (1999) "Nucleotides and Nucleosides", Chapter 8 in: Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology, Wiley: New York). Purine metabolism has been the subject of intensive research, and is essential to the normal functioning of the cell. Impaired purine metabolism in higher animals can cause severe disease, such as gout. Purine nucleotides are synthesized from ribose-5-phosphate, in a series of steps through the intermediate compound inosine-5'-phosphate (IMP), resulting in the production of guanosine-5'-monophosphate (GMP) or adenosine-5'-monophosphate (AMP), from which the triphosphate forms utilized as nucleotides are readily formed. These compounds are also utilized as energy stores, so their degradation provides energy for many different biochemical processes in the cell. Pyrimidine biosynthesis proceeds by the formation of uridine-5'-monophosphate (UMP) from ribose-5-phosphate. UMP, in turn, is converted to cytidine-5'-triphosphate (CTP). The deoxy- forms of all of these nucleotides are produced in a one step reduction reaction from the diphosphate ribose form of the nucleotide to the diphosphate deoxyribose form of the nucleotide. Upon phosphorylation, these molecules are able to participate in DNA synthesis.

D. Trehalose Metabolism and Uses

Trehalose consists of two glucose molecules, bound in α, α-1,1 linkage. It is commonly used in the food industry as a sweetener, an additive for dried or frozen foods, and in beverages. However, it also has applications in the pharmaceutical, cosmetics and biotechnology industries (see, for example, Nishimoto *et al.*, (1998) U.S. Patent No. 5,759,610; Singer, M.A. and Lindquist, S. (1998) *Trends Biotech.* 16: 460-467; Paiva, C.L.A. and Panek, A.D. (1996) *Biotech. Ann. Rev.* 2: 293-314; and Shiosaka, M. (1997) J. Japan 172: 97-102). Trehalose is produced by enzymes from many microorganisms and is naturally released into the surrounding medium, from which it can be collected using methods known in the art.

II. Maintenance of Homeostasis in *C. glutamicum* and Environmental Adaptation

The metabolic and other biochemical processes by which cells function are sensitive to environmental conditions such as temperature, pressure, solute concentration, and availability of oxygen. When one or more such environmental condition is perturbed or altered in a fashion that is incompatible with the normal functioning of these cellular processes, the cell must act to maintain an intracellular

environment which will permit them to occur despite the hostile extracellular environment. Gram positive bacterial cells, such as *C. glutamicum* cells, have a number of mechanisms by which internal homeostasis may be maintained despite unfavorable extracellular conditions. These include a cell wall, proteins which are able to degrade possibly toxic aromatic and aliphatic compounds, mechanisms of proteolysis whereby misfolded or misregulated proteins may be rapidly destroyed, and catalysts which permit intracellular reactions to occur which would not normally take place under the conditions optimal for bacterial growth.

Aside from merely surviving in a hostile environment, bacterial cells (*e.g. C. glutamicum* cells) are also frequently able to adapt such that they are able to take advantage of such conditions. For example, cells in an environment lacking desired carbon sources may be able to adapt to growth on a less-suitable carbon source. Also, cells may be able to utilize less desirable inorganic compounds when the commonly utilized ones are unavailable. *C. glutamicum* cells possess a number of genes which permit them to adapt to utilize inorganic and organic molecules which they would normally not encounter under optimal growth conditions as nutrients and precursors for metabolism. Aspects of cellular processes involved in homeostasis and adaptation are further explicated below.

A. Modification and Degradation of Aromatic and Aliphatic Compounds

Bacterial cells are routinely exposed to a variety of aromatic and aliphatic compounds in nature. Aromatic compounds are organic molecules having a cyclic ring structure, while aliphatic compounds are organic molecules having open chain structures rather than ring structures. Such compounds may arise as by-products of industrial processes (*e.g.*, benzene or toluene), but may also be produced by certain microorganisms (*e.g.*, alcohols). Many of these compounds are toxic to cells, particularly the aromatic compounds, which are highly reactive due to the high-energy ring structure. Thus, certain bacteria have developed mechanisms by which they are able to modify or degrade these compounds such that they are no longer hazardous to the cell. Cells may possess enzymes that are able to, for example, hydroxylate, isomerize, or methylate aromatic or aliphatic compounds such that they are either rendered less toxic, or such that the modified form is able to be processed by standard cellular waste and degradation pathways. Also, cells may possess enzymes which are able to specifically degrade one or more such potentially hazardous substance, thereby protecting the cell. Principles and examples of these types of modification and degradation processes in bacteria are described in several publications, *e.g.*, Sahm, H. (1999) "Procaryotes in Industrial Production" in Lengeler, J.W. *et al.*, eds.

Biology of the Procaryotes, Thieme Verlag: Stuttgart; and Schlegel, H.G. (1992) Allgemeine Mikrobiologie, Thieme: Stuttgart).

Aside from simply inactivating hazardous aromatic or aliphatic compounds, many bacteria have evolved to be able to utilize these compounds as carbon sources for continued metabolism when the preferred carbon sources of the cell are not available. For example, Pseudomonas strains able to utilize toluene, benzene, and 1,10-dichlorodecane as carbon sources are known (Chang, B.V. *et al.* (1997) *Chemosphere* 35(12): 2807-2815; Wischnak, C. *et al.* (1998) *Appl. Environ. Microbiol.* 64(9): 3507-3511; Churchill, S.A. *et al.* (1999) *Appl. Environ. Microbiol.* 65(2): 549-552). There are similar examples from many other bacterial species which are known in the art.

The ability of certain bacteria to modify or degrade aromatic and aliphatic compounds has begun to be exploited. Petroleum is a complex mixture of chemicals which includes aliphatic molecules and aromatic compounds. By applying bacteria having the ability to degrade or modify these toxic compounds to an oil spill, for example, it is possible to eliminate much of the environmental damage with high efficiency and low cost (see, for example, Smith, M.R. (1990) "The biodegradation of aromatic hydrocarbons by bacteria" *Biodegradation* 1(2-3): 191-206; and Suyama, T. *et al.* (1998) "Bacterial isolates degrading aliphatic polycarbonates," *FEMS Microbiol. Lett.* 161(2): 255-261).

B. Metabolism of Inorganic Compounds

Cells (*e.g.*, bacterial cells) contain large quantities of different molecules, such as water, inorganic ions, and organic substances (*e.g.*, proteins, sugars, and other macromolecules). The bulk of the mass of a typical cell consists of only 4 types of atoms: carbon, oxygen, hydrogen, and nitrogen. Although they represent a smaller percentage of the content of a cell, inorganic substances are equally as important to the proper functioning of the cell. Such molecules include phosphorous, sulfur, calcium, magnesium, iron, zinc, manganese, copper, molybdenum, tungsten, and cobalt. Many of these compounds are critical for the construction of important molecules, such as nucleotides (phosphorous) and amino acids (nitrogen and sulfur). Others of these inorganic ions serve as cofactors for enzymic reactions or contribute to osmotic pressure. All such molecules must be taken up by the bacterium from the surrounding environment.

For each of these inorganic compounds it is desirable for the bacterium to take up the form which can be most readily used by the standard metabolic machinery of the cell. However, the bacterium may encounter environments in which these preferred forms are not readily available. In order to survive under these circumstances, it is important for bacteria to have additional biochemical mechanisms which are able to convert less metabolically active but readily available forms of these inorganic compounds to ones which may be used

in cellular metabolism. Bacteria frequently possess a number of genes encoding enzymes for this purpose, which are not expressed unless the desired inorganic species are not available. Thus, these genes for the metabolism of various inorganic compounds serve as another tool which bacteria may use to adapt to suboptimal environmental conditions.

After carbon, the most important element in the cell is nitrogen. A typical bacterial cell contains between 12-15% nitrogen. It is a constituent of amino acids and nucleotides, as well as many other important molecules in the cell. Further, nitrogen may serve as a substitute for oxygen as a terminal electron acceptor in energy metabolism. Good sources of nitrogen include many organic and inorganic compounds, such ammonia gas or ammonia salts (*e.g.*, NH_4Cl, $(NH_4)_2SO_4$, or NH_4OH), nitrates, urea, amino acids, or complex nitrogen sources like corn steep liquor, soy bean flour, soy bean protein, yeast extract, meat extract, etc. Ammonia nitrogen is fixed by the action of particular enzymes: glutamate dehydrogenase, glutamine synthase, and glutamine-2-oxoglutarate aminotransferase. The transfer of amino-nitrogen from one organic molecule to another is accomplished by the aminotransferases, a class of enzymes which transfer one amino group from an alpha-amino acid to an alpha-keto acid. Nitrate may be reduced via nitrate reductase, nitrite reductase, and further redox enzymes until it is converted to molecular nitrogen or ammonia, which may be readily utilized by the cell in standard metabolic pathways.

Phosphorous is typically found intracellularly in both organic and inorganic forms, and may be taken up by the cell in either of these forms as well, though most microorganisms preferentially take up inorganic phosphate. The conversion of organic phosphate to a form which the cell can utilize requires the action of phosphatases (*e.g.*, phytases, which hydrolyze phyate-yielding phosphate and inositol derivatives). Phosphate is a key element in the synthesis of nucleic acids, and also has a significant role in cellular energy metabolism (*e.g.*, in the synthesis of ATP, ADP, and AMP).

Sulfur is a requirement for the synthesis of amino acids (*e.g.*, methionine and cysteine), vitamins (*e.g.*, thiamine, biotin, and lipoic acid) and iron sulfur proteins. Bacteria obtain sulfur primarily from inorganic sulfate, though thiosulfate, sulfite, and sulfide are also commonly utilized. Under conditions where these compounds may not be readily available, many bacteria express genes which enable them to utilize sulfonate compounds such as 2-aminosulfonate (taurine) (Kertesz, M.A. (1993) "Proteins induced by sulfate limitation in *Escherichia coli*, *Pseudomonas putida*, or *Staphylococcus aureus*." *J. Bacteriol.* 175: 1187-1190).

Other inorganic atoms, *e.g.*, metal or calcium ions, are also critical for the viability of cells. Iron, for example, plays a key role in redox reactions and is a cofactor of iron-sulfur proteins, heme proteins, and cytochromes. The uptake of iron into bacterial cells may be accomplished by the action of siderophores, chelating agents

which bind extracellular iron ions and translocate them to the interior of the cell. For reference on the metabolism of iron and other inorganic compounds, see: Lengeler *et al.* (1999) Biology of Prokaryotes, Thieme Verlag: Stuttgart; Neidhardt, F.C. *et al.*, eds. *Escherichia coli* and *Salmonella.* ASM Press: Washington, D.C.; Sonenshein, A.L. *et al.*, eds. (199?) *Bacillus subtilis* and Other Gram-Positive Bacteria, ASM Press: Washington, D.C.; Voet, D. and Voet, J.G. (1992) Biochemie, VCH: Weinheim; Brock, T.D. and Madigan, M.T. (1991) Biology of Microorganisms, 6th ed. Prentice Hall: Englewood Cliffs, p. 267-269; Rhodes, P.M. and Stanbury, P.F. Applied Microbial Physiology – A Practical Approach, Oxford Univ. Press: Oxford.

C. Enzymes and Proteolysis

The intracellular conditions for which bacteria such as *C. glutamicum* are optimized are frequently not conditions under which many biochemical reactions would normally take place. In order to make such reactions proceed under physiological conditions, cells utilize enzymes. Enzymes are proteinaceous biological catalysts, spatially orienting reacting molecules or providing a specialized environment such that the energy barrier to a biochemical reaction is lowered. Different enzymes catalyze different reactions, and each enzyme may be the subject of transcriptional, translational, or posttranslational regulation such that the reaction will only take place under appropriate conditions and at specified times. Enzymes may contribute to the degradation (*e.g.*, the proteases), synthesis (*e.g.*, the synthases), or modification (*e.g.*, transferases or isomerases) of compounds, all of which enable the production of necessary compounds within the cell. This, in turn, contributes to the maintenance of cellular homeostasis.

However, the fact that enzymes are optimized for activity under the physiological conditions at which the bacterium is most viable means that when environmental conditions are perturbed, there is a significant possibility that enzyme activity will also be perturbed. For example, changes in temperature may result in aberrantly folded proteins, and the same is true for changes of pH – protein folding is largely dependent on electrostatic and hydrophobic interactions of amino acids within the polypeptide chain, so any alteration to the charges on individual amino acids (as might be brought about by a change in cellular pH) may have a profound effect on the ability of the protein to correctly fold. Changes in temperature effectively change the amount of kinetic energy that the polypeptide molecule possesses, which affects the ability of the polypeptide to settle into a correctly folded, energetically stable configuration. Misfolded proteins may be harmful to the cell for two reasons. First, the aberrantly folded protein may have a similarly aberrant activity, or no activity

whatsoever. Second, misfolded proteins may lack the conformational regions necessary for proper regulation by other cellular systems and thus may continue to be active but in an uncontrolled fashion.

The cell has a mechanism by which misfolded enzymes and regulatory proteins may be rapidly destroyed before any damage occurs to the cell: proteolysis. Proteins such as those of the la/lon family and those of the Clp family specifically recognize and degrade misfolded proteins (see, *e.g.*, Sherman, M.Y., Goldberg, A.L. (1999) *EXS* 77: 57-78 and references therein and Porankiewicz J. (1999) *Molec. Microbiol.* 32(3): 449-58, and references therein; Neidhardt, F.C., *et al.* (1996) *E. coli* and *Salmonella*, ASM Press: Washington, D.C. and references therein; and Pritchard, G.G., and Coolbear, T. (1993) *FEMS Microbiol.* Rev. 12(1-3): 179-206 and references therein). These enzymes bind to misfolded or unfolded proteins and degrade them in an ATP-dependent manner. Proteolysis thus serves as an important mechanism employed by the cell to prevent damage to normal cellular functions upon environmental changes, and it further permits cells to survive under conditions and in environments which would otherwise be toxic due to misregulated and/or aberrant enzyme or regulatory activity.

Proteolysis also has important functions in the cell under optimal environmental conditions. Within normal metabolic processes, proteases aid in the hydrolysis of peptide bonds, in the catabolism of complex molecules to provide necessary degradation products, and in protein modification. Secreted proteases play an important role in the catabolism of external nutrients even prior to the entry of these compounds into the cell. Further, proteolytic activity itself may serve regulatory functions; sporulation in *B. subtilis* and cell cycle progression in *Caulobacter* spp. are known to be regulated by key proteolytic events in each of these species (Gottesman, S. (1999) *Curr. Opin. Microbiol.* 2(2): 142-147). Thus, proteolytic processes are key for cellular survival under both suboptimal and optimal environmental conditions, and contribute to the overall maintenance of homeostasis in cells.

D. Cell Wall Production and Rearrangements

While the biochemical machinery of the cell may be able to readily adapt to different and possibly unfavorable environments, cells still require a general mechanism by which they may be protected from the environment. For many bacteria, the cell wall affords such protection, and also plays roles in adhesion, cell growth and division, and transport of desired solutes and waste materials.

In order to function, cells require intracellular concentrations of metabolites and other molecules that are substantially higher than those of the surrounding media. Since these metabolites are largely prevented from leaving the cell due to the presence of the

hydrophobic membrane, the tendency of the system is for water molecules to enter the cell from the external medium such that the interior concentrations of solutes match the exterior concentrations. Water molecules are readily able to cross the cellular membrane, and this membrane is not able to withstand the resulting swelling and pressure, which may lead to osmotic lysis of the cell. The rigidity of the cell wall greatly improves the ability of the cell to tolerate these pressures, and offers a further barrier to the unwanted diffusion of these metabolites and desired solutes from the cell. Similarly, the cell wall also serves to prevent unwanted material from entering the cell.

The cell wall also participates in a number of other cellular processes, such as adhesion and cell growth and division. Due to the fact that the cell wall completely surrounds the cell, any interaction of the cell with its surroundings must be mediated by the cell wall. Thus, the cell wall must participate in any adherence of the cell to other cells and to desired surfaces. Further, the cell cannot grow or divide without concomitant changes in the cell wall. Since the protection that the wall affords requires its presence during growth, morphogenesis and multiplication, one of the key steps in cell division is cell wall synthesis within the cell such that a new cell divides from the old. Thus, frequently cell wall biosynthesis is regulated in tandem with cell growth and cell division (see, *e.g.*, Sonenshein, A.L. et al, eds. (1993) *Bacillus subtilis* and Other Gram-Positive Bacteria, ASM: Washington, D.C.).

The structure of the cell wall varies between gram-positive and gram-negative bacteria. However, in both types, the fundamental structural unit of the wall remains similar: an overlapping lattice of two polysaccharides, N-acetyl glucosamine (NAG) and N-acetyl muramic acid (NAM) which are cross-linked by amino acids (most commonly L-alanine, D-glutamate, diaminopimelic acid, and D-alanine), termed 'peptidoglycan'. The processes involved in the synthesis of the cell wall are known (see, *e.g.*, Michal, G., ed. (1999) Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology, Wiley: New York).

In gram-negative bacteria, the inner cellular membrane is coated by a single-layered peptidoglycan (approximately 10 nm thick), termed the murein-sacculus. This peptidoglycan structure is very rigid, and its structure determines the shape of the organism. The outer surface of the murein-sacculus is covered with an outer membrane, containing porins and other membrane proteins, phospholipids, and lipopolysaccharides. To maintain a tight association with the outer membrane, the gram-negative cell wall also has interspersed lipid molecules which serve to anchor it to the surrounding membrane.

In gram-positive bacteria, such as *Corynebacterium glutamicum*, the cytoplasmic membrane is covered by a multi-layered peptidoglycan, which ranges from 20-80 nm in thickness (see, *e.g.*, Lengeler *et al.* (1999) Biology of Prokaryotes Thieme Verlag: Stuttgart,

p. 913-918, p. 875-899, and p. 88-109 and references therein). The gram-positive cell wall also contains teichoic acid, a polymer of glycerol or ribitol linked through phosphate groups. Teichoic acid is also able to associate with amino acids, and forms covalent bonds with muramic acid. Also present in the cell wall may be lipoteichoic acids and teichuronic acids. If present, cellular surface structures such as flagella or capsules will be anchored in this layer as well.

III. Elements and Methods of the Invention

The present invention is based, at least in part, on the discovery of novel molecules, referred to herein as HA nucleic acid and protein molecules, which participate in the maintenance of homeostasis in *C. glutamicum*, or which perform a function involved in the adaptation of this microorganism to different environmental conditions. In one embodiment, the HA molecules participate in *C. glutamicum* cell wall biosynthesis or rearrangements, in the metabolism of inorganic compounds, in the modification or degradation of aromatic or aliphatic compounds, or have an enzymatic or proteolytic activity. In a preferred embodiment, the activity of the HA molecules of the present invention with regard to *C. glutamicum* cell wall biosynthesis or rearrangements, metabolism of inorganic compounds, modification or degradation of aromatic or aliphatic compounds, or enzymatic or proteolytic activity has an impact on the production of a desired fine chemical by this organism. In a particularly preferred embodiment, the HA molecules of the invention are modulated in activity, such that the *C. glutamicum* cellular processes in which the HA molecules participate (*e.g.*, *C. glutamicum* cell wall biosynthesis or rearrangements, metabolism of inorganic compounds, modification or degradation of aromatic or aliphatic compounds, or enzymatic or proteolytic activity) are also altered in activity, resulting either directly or indirectly in a modulation of the yield, production, and/or efficiency of production of a desired fine chemical by *C. glutamicum*.

The language, "HA protein" or "HA polypeptide" includes proteins which participate in a number of cellular processes related to *C. glutamicum* homeostasis or the ability of *C. glutamicum* cells to adapt to unfavorable environmental conditions. For example, an HA protein may be involved in *C. glutamicum* cell wall biosynthesis or rearrangements, in the metabolism of inorganic compounds in *C. glutamicum*, in the modification or degradation of aromatic or aliphatic compounds in *C. glutamicum*, or have a *C. glutamicum* enzymatic or proteolytic activity. Examples of HA proteins include those encoded by the HA genes set forth in Table 1 and Appendix A. The terms "HA gene" or "HA nucleic acid sequence" include nucleic acid sequences encoding an HA protein, which consist of a coding region and also corresponding untranslated 5' and

3' sequence regions. Examples of HA genes include those set forth in Table 1. The terms "production" or "productivity" are art-recognized and include the concentration of the fermentation product (for example, the desired fine chemical) formed within a given time and a given fermentation volume (*e.g.*, kg product per hour per liter). The term "efficiency of production" includes the time required for a particular level of production to be achieved (for example, how long it takes for the cell to attain a particular rate of output of a fine chemical). The term "yield" or "product/carbon yield" is art-recognized and includes the efficiency of the conversion of the carbon source into the product (*i.e.*, fine chemical). This is generally written as, for example, kg product per kg carbon source. By increasing the yield or production of the compound, the quantity of recovered molecules, or of useful recovered molecules of that compound in a given amount of culture over a given amount of time is increased. The terms "biosynthesis" or a "biosynthetic pathway" are art-recognized and include the synthesis of a compound, preferably an organic compound, by a cell from intermediate compounds in what may be a multistep and highly regulated process. The terms "degradation" or a "degradation pathway" are art-recognized and include the breakdown of a compound, preferably an organic compound, by a cell to degradation products (generally speaking, smaller or less complex molecules) in what may be a multistep and highly regulated process. The language "metabolism" is art-recognized and includes the totality of the biochemical reactions that take place in an organism. The metabolism of a particular compound, then, (*e.g.*, the metabolism of an amino acid such as glycine) comprises the overall biosynthetic, modification, and degradation pathways in the cell related to this compound. The term "homeostasis" is art-recognized and includes all of the mechanisms utilized by a cell to maintain a constant intracellular environment despite the prevailing extracellular environmental conditions. A non-limiting example of such processes is the utilization of a cell wall to prevent osmotic lysis due to high intracellular solute concentrations. The term "adaptation" or "adaptation to an environmental condition" is art-recognized and includes mechanisms utilized by the cell to render the cell able to survive under nonpreferred environmental conditions (generally speaking, those environmental conditions in which one or more favored nutrients are absent, or in which an environmental condition such as temperature, pH, osmolarity, oxygen percentage and the like fall outside of the optimal survival range of the cell). Many cells, including *C. glutamicum* cells, possess genes encoding proteins which are expressed under such environmental conditions and which permit continued growth in such suboptimal conditions.

In another embodiment, the HA molecules of the invention are capable of modulating the production of a desired molecule, such as a fine chemical, in a

microorganism such as *C. glutamicum*. There are a number of mechanisms by which the alteration of an HA protein of the invention may directly affect the yield, production, and/or efficiency of production of a fine chemical from a *C. glutamicum* strain incorporating such an altered protein. For example, by engineering enzymes which modify or degrade aromatic or aliphatic compounds such that these enzymes are increased or decreased in activity or number, it may be possible to modulate the production of one or more fine chemicals which are the modification or degradation products of these compounds. Similarly, enzymes involved in the metabolism of inorganic compounds provide key molecules (*e.g.* phosphorous, sulfur, and nitrogen molecules) for the biosynthesis of such fine chemicals as amino acids, vitamins, and nucleic acids. By altering the activity or number of these enzymes in *C. glutamicum*, it may be possible to increase the conversion of these inorganic compounds (or to use alternate inorganic compounds) to thus permit improved rates of incorporation of inorganic atoms into these fine chemicals. Genetic engineering of *C. glutamicum* enzymes involved in general cellular processes may also directly improve fine chemical production, since many of these enzymes directly modify fine chemicals (*e.g.*, amino acids) or the enzymes which are involved in fine chemical synthesis or secretion. Modulation of the activity or number of cellular proteases may also have a direct effect on fine chemical production, since many proteases may degrade fine chemicals or enzymes involved in fine chemical production or breakdown.

Further, the aforementioned enzymes which participate in aromatic/aliphatic compound modification or degradation, general biocatalysis, inorganic compound metabolism or proteolysis are each themselves fine chemicals, desirable for their activity in various *in vitro* industrial applications. By altering the number of copies of the gene for one or more of these enzymes in *C. glutamicum* it may be possible to increase the number of these proteins produced by the cell, thereby increasing the potential yield or efficiency of production of these proteins from large-scale *C. glutamicum* or related bacterial cultures.

The alteration of an HA protein of the invention may also indirectly affect the yield, production, and/or efficiency of production of a fine chemical from a *C. glutamicum* strain incorporating such an altered protein. For example, by modulating the activity and/or number of those proteins involved in the construction or rearrangement of the cell wall, it may be possible to modify the structure of the cell wall itself such that the cell is able to better withstand the mechanical and other stresses present during large-scale fermentative culture. Also, large-scale growth of *C. glutamicum* requires significant cell wall production. Modulation of the activity or number of cell wall biosynthetic or degradative enzymes may allow more rapid rates of

cell wall biosynthesis, which in turn may permit increased growth rates of this microorganism in culture and thereby increase the number of cells producing the desired fine chemical.

By modifying the HA enzymes of the invention, one may also indirectly impact the yield, production, or efficiency of production of one or more fine chemicals from *C. glutamicum*. For example, many of the general enzymes in *C. glutamicum* may have a significant impact on global cellular processes (*e.g.*, regulatory processes) which in turn have a significant effect on fine chemical metabolism. Similarly, proteases, enzymes which modify or degrade possibly toxic aromatic or aliphatic compounds, and enzymes which promote the metabolism of inorganic compounds all serve to increase the viability of *C. glutamicum*. The proteases aid in the selective removal of misfolded or misregulated proteins, such as those that might occur under the relatively stressful environmental conditions encountered during large-scale fermentor culture. By altering these proteins, it may be possible to further enhance this activity and to improve the viability of *C. glutamicum* in culture. The aromatic/aliphatic modification or degradation proteins not only serve to detoxify these waste compounds (which may be encountered as impurities in culture medium or as waste products from cells themselves), but also to permit the cells to utilize alternate carbon sources if the optimal carbon source is limiting in the culture. By increasing their number and/or activity, the survival of *C. glutamicum* cells in culture may be enhanced. The inorganic metabolism proteins of the invention supply the cell with inorganic molecules required for all protein and nucleotide (among others) synthesis, and thus are critical for the overall viability of the cell. An increase in the number of viable cells producing one or more desired fine chemicals in large-scale culture should result in a concomitant increase in the yield, production, and/or efficiency of production of the fine chemical in the culture.

The isolated nucleic acid sequences of the invention are contained within the genome of a *Corynebacterium glutamicum* strain available through the American Type Culture Collection, given designation ATCC 13032. The nucleotide sequence of the isolated *C. glutamicum* HA DNAs and the predicted amino acid sequences of the *C. glutamicum* HA proteins are shown in Appendices A and B, respectively. Computational analyses were performed which classified and/or identified these nucleotide sequences as sequences which encode proteins that participate in *C. glutamicum* cell wall biosynthesis or rearrangements, metabolism of inorganic compounds, modification or degradation of aromatic or aliphatic compounds, or that have a *C. glutamicum* enzymatic or proteolytic activity.

The present invention also pertains to proteins which have an amino acid sequence which is substantially homologous to an amino acid sequence of Appendix B.

As used herein, a protein which has an amino acid sequence which is substantially homologous to a selected amino acid sequence is least about 50% homologous to the selected amino acid sequence, *e.g.*, the entire selected amino acid sequence. A protein which has an amino acid sequence which is substantially homologous to a selected amino acid sequence can also be least about 50-60%, preferably at least about 60-70%, and more preferably at least about 70-80%, 80-90%, or 90-95%, and most preferably at least about 96%, 97%, 98%, 99% or more homologous to the selected amino acid sequence.

The HA protein or a biologically active portion or fragment thereof of the invention can participate in the maintenance of homeostasis in *C. glutamicum*, or can perform a function involved in the adaptation of this microorganism to different environmental conditions, or have one or more of the activities set forth in Table 1.

Various aspects of the invention are described in further detail in the following subsections.

A. Isolated Nucleic Acid Molecules

One aspect of the invention pertains to isolated nucleic acid molecules that encode HA polypeptides or biologically active portions thereof, as well as nucleic acid fragments sufficient for use as hybridization probes or primers for the identification or amplification of HA-encoding nucleic acid (*e.g.*, HA DNA). As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (*e.g.*, cDNA or genomic DNA) and RNA molecules (*e.g.*, mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. This term also encompasses untranslated sequence located at both the 3' and 5' ends of the coding region of the gene: at least about 100 nucleotides of sequence upstream from the 5' end of the coding region and at least about 20 nucleotides of sequence downstream from the 3'end of the coding region of the gene. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA. An "isolated" nucleic acid molecule is one which is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (*i.e.*, sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated HA nucleic acid molecule can contain less than about 5 kb, 4kb, 3kb, 2kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived (e.g, a *C. glutamicum* cell). Moreover, an "isolated" nucleic acid molecule, such as a DNA molecule, can be substantially free of other cellular material, or culture

medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized.

A nucleic acid molecule of the present invention, *e.g.*, a nucleic acid molecule having a nucleotide sequence of Appendix A, or a portion thereof, can be isolated using standard molecular biology techniques and the sequence information provided herein. For example, a *C. glutamicum* HA DNA can be isolated from a *C. glutamicum* library using all or portion of one of the sequences of Appendix A as a hybridization probe and standard hybridization techniques (*e.g.*, as described in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989). Moreover, a nucleic acid molecule encompassing all or a portion of one of the sequences of Appendix A can be isolated by the polymerase chain reaction using oligonucleotide primers designed based upon this sequence (*e.g.*, a nucleic acid molecule encompassing all or a portion of one of the sequences of Appendix A can be isolated by the polymerase chain reaction using oligonucleotide primers designed based upon this same sequence of Appendix A). For example, mRNA can be isolated from normal endothelial cells (*e.g.*, by the guanidinium-thiocyanate extraction procedure of Chirgwin *et al.* (1979) *Biochemistry* 18: 5294-5299) and DNA can be prepared using reverse transcriptase (*e.g.*, Moloney MLV reverse transcriptase, available from Gibco/BRL, Bethesda, MD; or AMV reverse transcriptase, available from Seikagaku America, Inc., St. Petersburg, FL). Synthetic oligonucleotide primers for polymerase chain reaction amplification can be designed based upon one of the nucleotide sequences shown in Appendix A. A nucleic acid of the invention can be amplified using cDNA or, alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to an HA nucleotide sequence can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer.

In a preferred embodiment, an isolated nucleic acid molecule of the invention comprises one of the nucleotide sequences shown in Appendix A. The sequences of Appendix A correspond to the *Corynebacterium glutamicum* HA DNAs of the invention. This DNA comprises sequences encoding HA proteins (*i.e.*, the "coding region", indicated in each sequence in Appendix A), as well as 5' untranslated sequences and 3' untranslated sequences, also indicated in Appendix A. Alternatively, the nucleic acid molecule can comprise only the coding region of any of the sequences in Appendix A.

For the purposes of this application, it will be understood that each of the sequences set forth in Appendix A has an identifying RXA, RXN, RXS, or RXC number having the designation "RXA", "RXN", "RXS", or "RXC" followed by 5 digits (*i.e.*, RXA02702, RXN02707, RXS02560, and RXC00110). Each of these sequences comprises up to three parts: a 5' upstream region, a coding region, and a downstream region. Each of these three regions is identified by the same RXA, RXN, RXS, or RXC designation to eliminate confusion. The recitation "one of the sequences in Appendix A", then, refers to any of the sequences in Appendix A, which may be distinguished by their differing RXA, RXN, RXS, or RXC designations. The coding region of each of these sequences is translated into a corresponding amino acid sequence, which is set forth in Appendix B. The sequences of Appendix B are identified by the same RXA, RXN, RXS, or RXC designations as Appendix A, such that they can be readily correlated. For example, the amino acid sequences in Appendix B designated RXA02702, RXN02707, RXS02560, and RXC00110 are translations of the coding regions of the nucleotide sequence of nucleic acid molecules RXA02702, RXN02707, RXS02560, and RXC00110, respectively, in Appendix A. Each of the RXA, RXN, RXS, and RXC nucleotide and amino acid sequences of the invention has also been assigned a SEQ ID NO, as indicated in Table 1.

Several of the genes of the invention are "F-designated genes". An F-designated gene includes those genes set forth in Table 1 which have an 'F' in front of the RXA, RXN, RXS, or RXC designation. For example, SEQ ID NO:1, designated, as indicated on Table 1, as "F RXA02702", is an F-designated gene, as are SEQ ID NOs: 9, 11, and 13 (designated on Table 1 as "F RXA02707", "F RXA02708", and "F RXA02709", respectively).

In one embodiment, the nucleic acid molecules of the present invention are not intended to include those compiled in Table 2. In the case of the dapD gene, a sequence for this gene was published in Wehrmann, A., *et al.* (1998) *J. Bacteriol.* 180(12): 3159-3165. However, the sequence obtained by the inventors of the present application is significantly longer than the published version. It is believed that the published version relied on an incorrect start codon, and thus represents only a fragment of the actual coding region.

In another preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule which is a complement of one of the nucleotide sequences shown in Appendix A, or a portion thereof. A nucleic acid molecule which is complementary to one of the nucleotide sequences shown in Appendix A is one which is sufficiently complementary to one of the nucleotide

sequences shown in Appendix A such that it can hybridize to one of the nucleotide sequences shown in Appendix A, thereby forming a stable duplex.

In still another preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleotide sequence which is at least about 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, or 60%, preferably at least about 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, or 70%%, more preferably at least about 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, or 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, or 90%, or 91%, 92%, 93%, 94%, and even more preferably at least about 95%, 96%, 97%, 98%, 99% or more homologous to a nucleotide sequence shown in Appendix A, or a portion thereof. Ranges and identity values intermediate to the above-recited ranges, (*e.g.*, 70-90% identical or 80-95% identical) are also intended to be encompassed by the present invention. For example, ranges of identity values using a combination of any of the above values recited as upper and/or lower limits are intended to be included. In an additional preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleotide sequence which hybridizes, *e.g.*, hybridizes under stringent conditions, to one of the nucleotide sequences shown in Appendix A, or a portion thereof.

Moreover, the nucleic acid molecule of the invention can comprise only a portion of the coding region of one of the sequences in Appendix A, for example a fragment which can be used as a probe or primer or a fragment encoding a biologically active portion of an HA protein. The nucleotide sequences determined from the cloning of the HA genes from *C. glutamicum* allows for the generation of probes and primers designed for use in identifying and/or cloning HA homologues in other cell types and organisms, as well as HA homologues from other *Corynebacteria* or related species. The probe/primer typically comprises substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, preferably about 25, more preferably about 40, 50 or 75 consecutive nucleotides of a sense strand of one of the sequences set forth in Appendix A, an anti-sense sequence of one of the sequences set forth in Appendix A, or naturally occurring mutants thereof. Primers based on a nucleotide sequence of Appendix A can be used in PCR reactions to clone HA homologues. Probes based on the HA nucleotide sequences can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In preferred embodiments, the probe further comprises a label group attached thereto, *e.g.* the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells which misexpress an HA protein, such as by measuring a level of an HA-encoding

nucleic acid in a sample of cells, *e.g.*, detecting HA mRNA levels or determining whether a genomic HA gene has been mutated or deleted.

In one embodiment, the nucleic acid molecule of the invention encodes a protein or portion thereof which includes an amino acid sequence which is sufficiently homologous to an amino acid sequence of Appendix B such that the protein or portion thereof maintains the ability to participate in the maintenance of homeostasis in *C. glutamicum*, or to perform a function involved in the adaptation of this microorganism to different environmental conditions. As used herein, the language "sufficiently homologous" refers to proteins or portions thereof which have amino acid sequences which include a minimum number of identical or equivalent (*e.g.*, an amino acid residue which has a similar side chain as an amino acid residue in one of the sequences of Appendix B) amino acid residues to an amino acid sequence of Appendix B such that the protein or portion thereof is able to participate in the maintenance of homeostasis in *C. glutamicum*, or to perform a function involved in the adaptation of this microorganism to different environmental conditions. Proteins involved in *C. glutamicum* cell wall biosynthesis or rearrangements, metabolism of inorganic compounds, modification or degradation of aromatic or aliphatic compounds, or that have a *C. glutamicum* enzymatic or proteolytic activity, as described herein, may play a role in the production and secretion of one or more fine chemicals. Examples of such activities are also described herein. Thus, "the function of an HA protein" contributes either directly or indirectly to the yield, production, and/or efficiency of production of one or more fine chemicals. Examples of HA protein activities are set forth in Table 1.

In another embodiment, the protein is at least about 50-60%, preferably at least about 60-70%, and more preferably at least about 70-80%, 80-90%, 90-95%, and most preferably at least about 96%, 97%, 98%, 99% or more homologous to an entire amino acid sequence of Appendix B.

Portions of proteins encoded by the HA nucleic acid molecules of the invention are preferably biologically active portions of one of the HA proteins. As used herein, the term "biologically active portion of an HA protein" is intended to include a portion, *e.g.*, a domain/motif, of an HA protein that can participate in the maintenance of homeostasis in *C. glutamicum*, or that can perform a function involved in the adaptation of this microorganism to different environmental conditions, or has an activity as set forth in Table 1. To determine whether an HA protein or a biologically active portion thereof can participate in *C. glutamicum* cell wall biosynthesis or rearrangements, metabolism of inorganic compounds, modification or degradation of aromatic or aliphatic compounds, or has a *C. glutamicum* enzymatic or proteolytic activity, an assay

of enzymatic activity may be performed. Such assay methods are well known to those of ordinary skill in the art, as detailed in Example 8 of the Exemplification.

Additional nucleic acid fragments encoding biologically active portions of an HA protein can be prepared by isolating a portion of one of the sequences in Appendix B, expressing the encoded portion of the HA protein or peptide (*e.g.*, by recombinant expression *in vitro*) and assessing the activity of the encoded portion of the HA protein or peptide.

The invention further encompasses nucleic acid molecules that differ from one of the nucleotide sequences shown in Appendix A (and portions thereof) due to degeneracy of the genetic code and thus encode the same HA protein as that encoded by the nucleotide sequences shown in Appendix A. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence shown in Appendix B. In a still further embodiment, the nucleic acid molecule of the invention encodes a full length *C. glutamicum* protein which is substantially homologous to an amino acid sequence of Appendix B (encoded by an open reading frame shown in Appendix A).

It will be understood by one of ordinary skill in the art that in one embodiment the sequences of the invention are not meant to include the sequences of the prior art, such as those Genbank sequences set forth in Tables 2 or 4 which were available prior to the present invention. In one embodiment, the invention includes nucleotide and amino acid sequences having a percent identity to a nucleotide or amino acid sequence of the invention which is greater than that of a sequence of the prior art (*e.g.*, a Genbank sequence (or the protein encoded by such a sequence) set forth in Tables 2 or 4). For example, the invention includes a nucleotide sequence which is greater than and/or at least 36% identical to the nucleotide sequence designated RXA00009 (SEQ ID NO:85), a nucleotide sequence which is greater than and/or at least 40% identical to the nucleotide sequence designated RXA00277 (SEQ ID NO:91), and a nucleotide sequence which is greater than and/or at least 43% identical to the nucleotide sequence designated RXA00499 (SEQ ID NO:173). One of ordinary skill in the art would be able to calculate the lower threshold of percent identity for any given sequence of the invention by examining the GAP-calculated percent identity scores set forth in Table 4 for each of the three top hits for the given sequence, and by subtracting the highest GAP-calculated percent identity from 100 percent. One of ordinary skill in the art will also appreciate that nucleic acid and amino acid sequences having percent identities greater than the lower threshold so calculated (*e.g.*, at least 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, or 60%, preferably at least about 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, or 70%, more preferably at least about 71%, 72%, 73%, 74%, 75%,

76%, 77%, 78%, 79%, or 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, or 90%, or 91%, 92%, 93%, 94%, and even more preferably at least about 95%, 96%, 97%, 98%, 99% or more identical) are also encompassed by the invention.

In addition to the *C. glutamicum* HA nucleotide sequences shown in Appendix A, it will be appreciated by those of ordinary skill in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of HA proteins may exist within a population (*e.g.*, the *C. glutamicum* population). Such genetic polymorphism in the HA gene may exist among individuals within a population due to natural variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame encoding an HA protein, preferably a *C. glutamicum* HA protein. Such natural variations can typically result in 1-5% variance in the nucleotide sequence of the HA gene. Any and all such nucleotide variations and resulting amino acid polymorphisms in HA that are the result of natural variation and that do not alter the functional activity of HA proteins are intended to be within the scope of the invention.

Nucleic acid molecules corresponding to natural variants and non-*C. glutamicum* homologues of the *C. glutamicum* HA DNA of the invention can be isolated based on their homology to the *C. glutamicum* HA nucleic acid disclosed herein using the *C. glutamicum* DNA, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions. Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 15 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising a nucleotide sequence of Appendix A. In other embodiments, the nucleic acid is at least 30, 50, 100, 250 or more nucleotides in length. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% homologous to each other typically remain hybridized to each other. Preferably, the conditions are such that sequences at least about 65%, more preferably at least about 70%, and even more preferably at least about 75% or more homologous to each other typically remain hybridized to each other. Such stringent conditions are known to those of ordinary skill in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. A preferred, non-limiting example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65°C. Preferably, an isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to a sequence of Appendix A corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule

refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (*e.g.*, encodes a natural protein). In one embodiment, the nucleic acid encodes a natural *C. glutamicum* HA protein.

In addition to naturally-occurring variants of the HA sequence that may exist in the population, one of ordinary skill in the art will further appreciate that changes can be introduced by mutation into a nucleotide sequence of Appendix A, thereby leading to changes in the amino acid sequence of the encoded HA protein, without altering the functional ability of the HA protein. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in a sequence of Appendix A. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of one of the HA proteins (Appendix B) without altering the activity of said HA protein, whereas an "essential" amino acid residue is required for HA protein activity. Other amino acid residues, however, (*e.g.*, those that are not conserved or only semi-conserved in the domain having HA activity) may not be essential for activity and thus are likely to be amenable to alteration without altering HA activity.

Accordingly, another aspect of the invention pertains to nucleic acid molecules encoding HA proteins that contain changes in amino acid residues that are not essential for HA activity. Such HA proteins differ in amino acid sequence from a sequence contained in Appendix B yet retain at least one of the HA activities described herein. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 50% homologous to an amino acid sequence of Appendix B and is capable of participating in the maintenance of homeostasis in *C. glutamicum*, or of performing a function involved in the adaptation of this microorganism to different environmental conditions, or has one or more of the activities set forth in Table 1. Preferably, the protein encoded by the nucleic acid molecule is at least about 50-60% homologous to one of the sequences in Appendix B, more preferably at least about 60-70% homologous to one of the sequences in Appendix B, even more preferably at least about 70-80%, 80-90%, 90-95% homologous to one of the sequences in Appendix B, and most preferably at least about 96%, 97%, 98%, or 99% homologous to one of the sequences in Appendix B.

To determine the percent homology of two amino acid sequences (*e.g.*, one of the sequences of Appendix B and a mutant form thereof) or of two nucleic acids, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in the sequence of one protein or nucleic acid for optimal alignment with the other protein or nucleic acid). The amino acid residues or nucleotides at corresponding amino acid

positions or nucleotide positions are then compared. When a position in one sequence (*e.g.*, one of the sequences of Appendix B) is occupied by the same amino acid residue or nucleotide as the corresponding position in the other sequence (*e.g.*, a mutant form of the sequence selected from Appendix B), then the molecules are homologous at that position (*i.e.*, as used herein amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity"). The percent homology between the two sequences is a function of the number of identical positions shared by the sequences (*i.e.*, % homology = # of identical positions/total # of positions x 100).

An isolated nucleic acid molecule encoding an HA protein homologous to a protein sequence of Appendix B can be created by introducing one or more nucleotide substitutions, additions or deletions into a nucleotide sequence of Appendix A such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein. Mutations can be introduced into one of the sequences of Appendix A by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in an HA protein is preferably replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of an HA coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for an HA activity described herein to identify mutants that retain HA activity. Following mutagenesis of one of the sequences of Appendix A, the encoded protein can be expressed recombinantly and the activity of the protein can be determined using, for example, assays described herein (see Example 8 of the Exemplification).

In addition to the nucleic acid molecules encoding HA proteins described above, another aspect of the invention pertains to isolated nucleic acid molecules which are antisense thereto. An "antisense" nucleic acid comprises a nucleotide sequence which is complementary to a "sense" nucleic acid encoding a protein, *e.g.*, complementary to the

coding strand of a double-stranded DNA molecule or complementary to an mRNA sequence. Accordingly, an antisense nucleic acid can hydrogen bond to a sense nucleic acid. The antisense nucleic acid can be complementary to an entire HA coding strand, or to only a portion thereof. In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding an HA protein. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues (*e.g.*, the entire coding region of SEQ ID NO. 1 (RXA02702) comprises nucleotides 1 to1458). In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding HA. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding HA disclosed herein (*e.g.*, the sequences set forth in Appendix A), antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of HA mRNA, but more preferably is an oligonucleotide which is antisense to only a portion of the coding or noncoding region of HA mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of HA mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine,

2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a cell or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding an HA protein to thereby inhibit expression of the protein, *e.g.*, by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule which binds to DNA duplexes, through specific interactions in the major groove of the double helix. The antisense molecule can be modified such that it specifically binds to a receptor or an antigen expressed on a selected cell surface, *e.g.*, by linking the antisense nucleic acid molecule to a peptide or an antibody which binds to a cell surface receptor or antigen. The antisense nucleic acid molecule can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong prokaryotic, viral, or eukaryotic promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an α-anomeric nucleic acid molecule. An α-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β-units, the strands run parallel to each other (Gaultier *et al.* (1987) *Nucleic Acids. Res.* 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue *et al.* (1987) *Nucleic Acids Res.* 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue *et al.* (1987) *FEBS Lett.* 215:327-330).

In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity which are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes (described in Haselhoff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave HA mRNA transcripts to thereby inhibit translation of HA mRNA. A ribozyme having specificity for an HA-encoding nucleic acid can be designed based upon the nucleotide sequence of an HA DNA molecule disclosed herein (*i.e.*, SEQ ID NO. 3 (RXA02705) Appendix A). For example, a derivative of a *Tetrahymena* L-19

IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in an HA-encoding mRNA. See, *e.g.*, Cech *et al.* U.S. Patent No. 4,987,071 and Cech *et al.* U.S. Patent No. 5,116,742. Alternatively, HA mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, *e.g.*, Bartel, D. and Szostak, J.W. (1993) *Science* 261:1411-1418.

Alternatively, HA gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of an HA nucleotide sequence (*e.g.*, an HA promoter and/or enhancers) to form triple helical structures that prevent transcription of an HA gene in target cells. See generally, Helene, C. (1991) *Anticancer Drug Des.* 6(6):569-84; Helene, C. *et al.* (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher, L.J. (1992) *Bioassays* 14(12):807-15.

B. Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding an HA protein (or a portion thereof). As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which are

operatively linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel; *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Regulatory sequences include those which direct constitutive expression of a nucleotide sequence in many types of host cell and those which direct expression of the nucleotide sequence only in certain host cells. Preferred regulatory sequences are, for example, promoters such as cos-, tac-, trp-, tet-, trp-tet-, lpp-, lac-, lpp-lac-, $lacI^q$-, T7-, T5-, T3-, gal-, trc-, ara-, SP6-, arny, SPO2, λ-P_R- or λ P_L, which are used preferably in bacteria. Additional regulatory sequences are, for example, promoters from yeasts and fungi, such as ADC1, MFα, AC, P-60, CYC1, GAPDH, TEF, rp28, ADH, promoters from plants such as CaMV/35S, SSU, OCS, lib4, usp, STLS1, B33, nos or ubiquitin- or phaseolin-promoters. It is also possible to use artificial promoters. It will be appreciated by those of ordinary skill in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (*e.g.*, HA proteins, mutant forms of HA proteins, fusion proteins, etc.).

The recombinant expression vectors of the invention can be designed for expression of HA proteins in prokaryotic or eukaryotic cells. For example, HA genes can be expressed in bacterial cells such as *C. glutamicum*, insect cells (using baculovirus expression vectors), yeast and other fungal cells (see Romanos, M.A. *et al.* (1992) "Foreign gene expression in yeast: a review", *Yeast* 8: 423-488; van den Hondel, C.A.M.J.J. *et al.* (1991) "Heterologous gene expression in filamentous fungi" in: More Gene Manipulations in Fungi, J.W. Bennet & L.L. Lasure, eds., p. 396-428: Academic Press: San Diego; and van den Hondel, C.A.M.J.J. & Punt, P.J. (1991) "Gene transfer systems and vector development for filamentous fungi, in: Applied Molecular Genetics of Fungi, Peberdy, J.F. *et al.*, eds., p. 1-28, Cambridge University Press: Cambridge), algae and multicellular plant cells (see Schmidt, R. and Willmitzer, L. (1988) High efficiency *Agrobacterium tumefaciens* –mediated transformation of *Arabidopsis thaliana* leaf and cotyledon explants" *Plant Cell Rep.*: 583-586), or mammalian cells. Suitable host cells are discussed further in Goeddel, *Gene Expression Technology:*

Methods in Enzymology 185, Academic Press, San Diego, CA (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein but also to the C-terminus or fused within suitable regions in the proteins. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase.

Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith, D.B. and Johnson, K.S. (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein. In one embodiment, the coding sequence of the HA protein is cloned into a pGEX expression vector to create a vector encoding a fusion protein comprising, from the N-terminus to the C-terminus, GST-thrombin cleavage site-X protein. The fusion protein can be purified by affinity chromatography using glutathione-agarose resin. Recombinant HA protein unfused to GST can be recovered by cleavage of the fusion protein with thrombin.

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann *et al.*, (1988) *Gene* 69:301-315) pLG338, pACYC184, pBR322, pUC18, pUC19, pKC30, pRep4, pHS1, pHS2, pPLc236, pMBL24, pLG200, pUR290, pIN-III113-B1, λgt11, pBdCl, and pET 11d (Studier *et al.*, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 60-89 ; and Pouwels *et al.*, eds. (1985) Cloning Vectors. Elsevier: New York IBSN 0 444 904018). Target gene expression from the pTrc vector relies on host RNA polymerase transcription from a hybrid trp-lac fusion promoter. Target gene expression from the pET 11d vector relies on transcription from a T7 gn10-lac fusion promoter mediated by a coexpressed viral RNA polymerase (T7 gn1). This viral polymerase is supplied by host strains BL21(DE3) or HMS174(DE3) from a resident λ prophage harboring a T7 gn1 gene under the transcriptional control of the lacUV 5 promoter. For transformation

of other varieties of bacteria, appropriate vectors may be selected. For example, the plasmids pIJ101, pIJ364, pIJ702 and pIJ361 are known to be useful in transforming Streptomyces, while plasmids pUB110, pC194, or pBD214 are suited for transformation of Bacillus species. Several plasmids of use in the transfer of genetic information into Corynebacterium include pHM1519, pBL1, pSA77, or pAJ667 (Pouwels *et al.*, eds. (1985) Cloning Vectors. Elsevier: New York IBSN 0 444 904018).

One strategy to maximize recombinant protein expression is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in the bacterium chosen for expression, such as *C. glutamicum* (Wada *et al.* (1992) *Nucleic Acids Res.* 20: 2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the HA protein expression vector is a yeast expression vector. Examples of vectors for expression in yeast *S. cerevisiae* include pYepSec1 (Baldari, *et al.*, (1987) *Embo J.* 6:229-234), 2 μ, pAG-1, Yep6, Yep13, pEMBLYe23, pMFa (Kurjan and Herskowitz, (1982) *Cell* 30:933-943), pJRY88 (Schultz *et al.*, (1987) *Gene* 54:113-123), and pYES2 (Invitrogen Corporation, San Diego, CA). Vectors and methods for the construction of vectors appropriate for use in other fungi, such as the filamentous fungi, include those detailed in: van den Hondel, C.A.M.J.J. & Punt, P.J. (1991) "Gene transfer systems and vector development for filamentous fungi, in: Applied Molecular Genetics of Fungi, J.F. Peberdy, *et al.*, eds., p. 1-28, Cambridge University Press: Cambridge, and Pouwels *et al.*, eds. (1985) Cloning Vectors. Elsevier: New York (IBSN 0 444 904018).

Alternatively, the HA proteins of the invention can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (*e.g.*, Sf 9 cells) include the pAc series (Smith *et al.* (1983) *Mol. Cell Biol.* 3:2156-2165) and the pVL series (Lucklow and Summers (1989) *Virology* 170:31-39).

In another embodiment, the HA proteins of the invention may be expressed in unicellular plant cells (such as algae) or in plant cells from higher plants (*e.g.*, the spermatophytes, such as crop plants). Examples of plant expression vectors include those detailed in: Becker, D., Kemper, E., Schell, J. and Masterson, R. (1992) "New plant binary vectors with selectable markers located proximal to the left border", *Plant Mol. Biol.* 20: 1195-1197; and Bevan, M.W. (1984) "Binary *Agrobacterium* vectors for

plant transformation", *Nucl. Acid. Res.* 12: 8711-8721, and include pLGV23, pGHlac+, pBIN19, pAK2004, and pDH51 (Pouwels *et al.*, eds. (1985) Cloning Vectors. Elsevier: New York IBSN 0 444 904018).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, B. (1987) *Nature* 329:840) and pMT2PC (Kaufman *et al.* (1987) *EMBO J.* 6:187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see chapters 16 and 17 of Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (*e.g.*, tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert *et al.* (1987) *Genes Dev.* 1:268-277), lymphoid-specific promoters (Calame and Eaton (1988) *Adv. Immunol.* 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore (1989) *EMBO J.* 8:729-733) and immunoglobulins (Banerji *et al.* (1983) *Cell* 33:729-740; Queen and Baltimore (1983) *Cell* 33:741-748), neuron-specific promoters (*e.g.*, the neurofilament promoter; Byrne and Ruddle (1989) *PNAS* 86:5473-5477), pancreas-specific promoters (Edlund *et al.* (1985) *Science* 230:912-916), and mammary gland-specific promoters (*e.g.*, milk whey promoter; U.S. Patent No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, for example the murine hox promoters (Kessel and Gruss (1990) *Science* 249:374-379) and the α-fetoprotein promoter (Campes and Tilghman (1989) *Genes Dev.* 3:537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively linked to a regulatory sequence in a manner which allows for expression (by transcription of the DNA molecule) of an RNA molecule which is antisense to HA mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the continuous expression of the antisense RNA molecule in a variety of cell types, for

instance viral promoters and/or enhancers, or regulatory sequences can be chosen which direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see Weintraub, H. *et al.*, Antisense RNA as a molecular tool for genetic analysis, *Reviews - Trends in Genetics*, Vol. 1(1) (1986).

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, an HA protein can be expressed in bacterial cells such as *C. glutamicum*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those of ordinary skill in the art. Microorganisms related to *Corynebacterium glutamicum* which may be conveniently used as host cells for the nucleic acid and protein molecules of the invention are set forth in Table 3.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection", "conjugation" and "transduction" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.*, linear DNA or RNA (*e.g.*, a linearized vector or a gene construct alone without a vector) or nucleic acid in the form of a vector (*e.g.*, a plasmid, phage, phasmid, phagemid, transposon or other DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, natural competence, chemical-mediated transfer, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.* (*Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these

integrants, a gene that encodes a selectable marker (*e.g.*, resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those which confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding an HA protein or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by, for example, drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

To create a homologous recombinant microorganism, a vector is prepared which contains at least a portion of an HA gene into which a deletion, addition or substitution has been introduced to thereby alter, *e.g.*, functionally disrupt, the HA gene. Preferably, this HA gene is a *Corynebacterium glutamicum* HA gene, but it can be a homologue from a related bacterium or even from a mammalian, yeast, or insect source. In a preferred embodiment, the vector is designed such that, upon homologous recombination, the endogenous HA gene is functionally disrupted (*i.e.*, no longer encodes a functional protein; also referred to as a "knock out" vector). Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous HA gene is mutated or otherwise altered but still encodes functional protein (*e.g.*, the upstream regulatory region can be altered to thereby alter the expression of the endogenous HA protein). In the homologous recombination vector, the altered portion of the HA gene is flanked at its 5' and 3' ends by additional nucleic acid of the HA gene to allow for homologous recombination to occur between the exogenous HA gene carried by the vector and an endogenous HA gene in a microorganism. The additional flanking HA nucleic acid is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5' and 3' ends) are included in the vector (see *e.g.*, Thomas, K.R., and Capecchi, M.R. (1987) Cell 51: 503 for a description of homologous recombination vectors). The vector is introduced into a microorganism (*e.g.*, by electroporation) and cells in which the introduced HA gene has homologously recombined with the endogenous HA gene are selected, using art-known techniques.

In another embodiment, recombinant microorganisms can be produced which contain selected systems which allow for regulated expression of the introduced gene. For example, inclusion of an HA gene on a vector placing it under control of the lac operon permits expression of the HA gene only in the presence of IPTG. Such regulatory systems are well known in the art.

In another embodiment, an endogenous HA gene in a host cell is disrupted (*e.g.*, by homologous recombination or other genetic means known in the art) such that

expression of its protein product does not occur. In another embodiment, an endogenous or introduced HA gene in a host cell has been altered by one or more point mutations, deletions, or inversions, but still encodes a functional HA protein. In still another embodiment, one or more of the regulatory regions (*e.g.*, a promoter, repressor, or inducer) of an HA gene in a microorganism has been altered (*e.g.*, by deletion, truncation, inversion, or point mutation) such that the expression of the HA gene is modulated. One of ordinary skill in the art will appreciate that host cells containing more than one of the described HA gene and protein modifications may be readily produced using the methods of the invention, and are meant to be included in the present invention.

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.*, express) an HA protein. Accordingly, the invention further provides methods for producing HA proteins using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding an HA protein has been introduced, or into which genome has been introduced a gene encoding a wild-type or altered HA protein) in a suitable medium until HA protein is produced. In another embodiment, the method further comprises isolating HA proteins from the medium or the host cell.

C. Isolated HA Proteins

Another aspect of the invention pertains to isolated HA proteins, and biologically active portions thereof. An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material when produced by recombinant DNA techniques, or chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of HA protein in which the protein is separated from cellular components of the cells in which it is naturally or recombinantly produced. In one embodiment, the language "substantially free of cellular material" includes preparations of HA protein having less than about 30% (by dry weight) of non-HA protein (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-HA protein, still more preferably less than about 10% of non-HA protein, and most preferably less than about 5% non-HA protein. When the HA protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, *i.e.*, culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation. The language "substantially free of chemical precursors or other chemicals" includes

preparations of HA protein in which the protein is separated from chemical precursors or other chemicals which are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of HA protein having less than about 30% (by dry weight) of chemical precursors or non-HA chemicals, more preferably less than about 20% chemical precursors or non-HA chemicals, still more preferably less than about 10% chemical precursors or non-HA chemicals, and most preferably less than about 5% chemical precursors or non-HA chemicals. In preferred embodiments, isolated proteins or biologically active portions thereof lack contaminating proteins from the same organism from which the HA protein is derived. Typically, such proteins are produced by recombinant expression of, for example, a *C. glutamicum* HA protein in a microorganism such as *C. glutamicum.*

An isolated HA protein or a portion thereof of the invention can participate in the repair or recombination of DNA, in the transposition of genetic material, in gene expression (*i.e.*, the processes of transcription or translation), in protein folding, or in protein secretion in *Corynebacterium glutamicum*, or has one or more of the activities set forth in Table 1. In preferred embodiments, the protein or portion thereof comprises an amino acid sequence which is sufficiently homologous to an amino acid sequence of Appendix B such that the protein or portion thereof maintains the ability to participate in the maintenance of homeostasis in *C. glutamicum*, or to perform a function involved in the adaptation of this microorganism to different environmental conditions. The portion of the protein is preferably a biologically active portion as described herein. In another preferred embodiment, an HA protein of the invention has an amino acid sequence shown in Appendix B. In yet another preferred embodiment, the HA protein has an amino acid sequence which is encoded by a nucleotide sequence which hybridizes, *e.g.*, hybridizes under stringent conditions, to a nucleotide sequence of Appendix A. In still another preferred embodiment, the HA protein has an amino acid sequence which is encoded by a nucleotide sequence that is at least about 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, or 60%, preferably at least about 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, or 70%, more preferably at least about 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, or 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, or 90%, or 91%, 92%, 93%, 94%, and even more preferably at least about 95%, 96%, 97%, 98%, 99% % or more homologous to one of the nucleic acid sequences of Appendix A, or a portion thereof. Ranges and identity values intermediate to the above-recited values, (*e.g.*, 70-90% identical or 80-95% identical) are also intended to be encompassed by the present invention. For example, ranges of identity values using a combination of any of the above values recited as upper and/or lower limits are

intended to be included. The preferred HA proteins of the present invention also preferably possess at least one of the HA activities described herein. For example, a preferred HA protein of the present invention includes an amino acid sequence encoded by a nucleotide sequence which hybridizes, *e.g.*, hybridizes under stringent conditions, to a nucleotide sequence of Appendix A, and which can participate in the maintenance of homeostasis in *C. glutamicum*, or can perform a function involved in the adaptation of this microorganism to different environmental conditions, or which has one or more of the activities set forth in Table 1.

In other embodiments, the HA protein is substantially homologous to an amino acid sequence of Appendix B and retains the functional activity of the protein of one of the sequences of Appendix B yet differs in amino acid sequence due to natural variation or mutagenesis, as described in detail in subsection I above. Accordingly, in another embodiment, the HA protein is a protein which comprises an amino acid sequence which is at least about 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, or 60%, preferably at least about 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, or 70%, more preferably at least about 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, or 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, or 90%, or 91%, 92%, 93%, 94%, and even more preferably at least about 95%, 96%, 97%, 98%, 99% or more homologous to an entire amino acid sequence of Appendix B and which has at least one of the HA activities described herein. Ranges and identity values intermediate to the above-recited values, (*e.g.*, 70-90% identical or 80-95% identical) are also intended to be encompassed by the present invention. For example, ranges of identity values using a combination of any of the above values recited as upper and/or lower limits are intended to be included. In another embodiment, the invention pertains to a full length *C. glutamicum* protein which is substantially homologous to an entire amino acid sequence of Appendix B.

Biologically active portions of an HA protein include peptides comprising amino acid sequences derived from the amino acid sequence of an HA protein, *e.g.*, the an amino acid sequence shown in Appendix B or the amino acid sequence of a protein homologous to an HA protein, which include fewer amino acids than a full length HA protein or the full length protein which is homologous to an HA protein, and exhibit at least one activity of an HA protein. Typically, biologically active portions (peptides, *e.g.*, peptides which are, for example, 5, 10, 15, 20, 30, 35, 36, 37, 38, 39, 40, 50, 100 or more amino acids in length) comprise a domain or motif with at least one activity of an HA protein. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the activities described herein. Preferably, the biologically active portions of an

HA protein include one or more selected domains/motifs or portions thereof having biological activity.

HA proteins are preferably produced by recombinant DNA techniques. For example, a nucleic acid molecule encoding the protein is cloned into an expression vector (as described above), the expression vector is introduced into a host cell (as described above) and the HA protein is expressed in the host cell. The HA protein can then be isolated from the cells by an appropriate purification scheme using standard protein purification techniques. Alternative to recombinant expression, an HA protein, polypeptide, or peptide can be synthesized chemically using standard peptide synthesis techniques. Moreover, native HA protein can be isolated from cells (*e.g.*, endothelial cells), for example using an anti-HA antibody, which can be produced by standard techniques utilizing an HA protein or fragment thereof of this invention.

The invention also provides HA chimeric or fusion proteins. As used herein, an HA "chimeric protein" or "fusion protein" comprises an HA polypeptide operatively linked to a non-HA polypeptide. An "HA polypeptide" refers to a polypeptide having an amino acid sequence corresponding to an HA protein, whereas a "non-HA polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein which is not substantially homologous to the HA protein, *e.g.*, a protein which is different from the HA protein and which is derived from the same or a different organism. Within the fusion protein, the term "operatively linked" is intended to indicate that the HA polypeptide and the non-HA polypeptide are fused in-frame to each other. The non-HA polypeptide can be fused to the N-terminus or C-terminus of the HA polypeptide. For example, in one embodiment the fusion protein is a GST-HA fusion protein in which the HA sequences are fused to the C-terminus of the GST sequences. Such fusion proteins can facilitate the purification of recombinant HA proteins. In another embodiment, the fusion protein is an HA protein containing a heterologous signal sequence at its N-terminus. In certain host cells (*e.g.*, mammalian host cells), expression and/or secretion of an HA protein can be increased through use of a heterologous signal sequence.

Preferably, an HA chimeric or fusion protein of the invention is produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, for example by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor

primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for example, *Current Protocols in Molecular Biology*, eds. Ausubel *et al.* John Wiley & Sons: 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). An HA-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the HA protein.

Homologues of the HA protein can be generated by mutagenesis, *e.g.*, discrete point mutation or truncation of the HA protein. As used herein, the term "homologue" refers to a variant form of the HA protein which acts as an agonist or antagonist of the activity of the HA protein. An agonist of the HA protein can retain substantially the same, or a subset, of the biological activities of the HA protein. An antagonist of the HA protein can inhibit one or more of the activities of the naturally occurring form of the HA protein, by, for example, competitively binding to a downstream or upstream member of a biochemical cascade which includes the HA protein, by binding to a target molecule with which the HA protein interacts, such that no functional interaction is possible, or by binding directly to the HA protein and inhibiting its normal activity.

In an alternative embodiment, homologues of the HA protein can be identified by screening combinatorial libraries of mutants, *e.g.*, truncation mutants, of the HA protein for HA protein agonist or antagonist activity. In one embodiment, a variegated library of HA variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of HA variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential HA sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (*e.g.*, for phage display) containing the set of HA sequences therein. There are a variety of methods which can be used to produce libraries of potential HA homologues from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential HA sequences. Methods for synthesizing degenerate oligonucleotides are known in the art (see, *e.g.*, Narang, S.A. (1983) *Tetrahedron* 39:3; Itakura *et al.* (1984) *Annu. Rev. Biochem.* 53:323; Itakura *et al.* (1984) *Science* 198:1056; Ike *et al.* (1983) *Nucleic Acid Res.* 11:477.

In addition, libraries of fragments of the HA protein coding can be used to generate a variegated population of HA fragments for screening and subsequent

selection of homologues of an HA protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of an HA coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA which can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes N-terminal, C-terminal and internal fragments of various sizes of the HA protein.

Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of HA homologues. The most widely used techniques, which are amenable to high through-put analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a new technique which enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify HA homologues (Arkin and Yourvan (1992) *PNAS* *89*:7811-7815; Delgrave *et al.* (1993) *Protein Engineering* 6(3):327-331).

In another embodiment, cell based assays can be exploited to analyze a variegated HA library, using methods well known in the art.

D. Uses and Methods of the Invention

The nucleic acid molecules, proteins, protein homologues, fusion proteins, primers, vectors, and host cells described herein can be used in one or more of the following methods: identification of *C. glutamicum* and related organisms; mapping of genomes of organisms related to *C. glutamicum*; identification and localization of *C. glutamicum* sequences of interest; evolutionary studies; determination of HA protein regions required for function; modulation of an HA protein activity; modulation of the metabolism of one or more inorganic compounds; modulation of the modification or degradation of one or more aromatic or aliphatic compounds; modulation of cell wall synthesis or rearrangements; modulation of enzyme activity or proteolysis; and modulation of cellular production of a desired compound, such as a fine chemical.

The HA nucleic acid molecules of the invention have a variety of uses. First,

they may be used to identify an organism as being *Corynebacterium glutamicum* or a close relative thereof. Also, they may be used to identify the presence of *C. glutamicum* or a relative thereof in a mixed population of microorganisms. The invention provides the nucleic acid sequences of a number of *C. glutamicum* genes; by probing the extracted genomic DNA of a culture of a unique or mixed population of microorganisms under stringent conditions with a probe spanning a region of a *C. glutamicum* gene which is unique to this organism, one can ascertain whether this organism is present. Although *Corynebacterium glutamicum* itself is nonpathogenic, it is related to pathogenic species, such as *Corynebacterium diphtheriae*. *Corynebacterium diphtheriae* is the causative agent of diphtheria, a rapidly developing, acute, febrile infection which involves both local and systemic pathology. In this disease, a local lesion develops in the upper respiratory tract and involves necrotic injury to epithelial cells; the bacilli secrete toxin which is disseminated through this lesion to distal susceptible tissues of the body. Degenerative changes brought about by the inhibition of protein synthesis in these tissues, which include heart, muscle, peripheral nerves, adrenals, kidneys, liver and spleen, result in the systemic pathology of the disease. Diphtheria continues to have high incidence in many parts of the world, including Africa, Asia, Eastern Europe and the independent states of the former Soviet Union. An ongoing epidemic of diphtheria in the latter two regions has resulted in at least 5,000 deaths since 1990.

In one embodiment, the invention provides a method of identifying the presence or activity of *Cornyebacterium diphtheriae* in a subject. This method includes detection of one or more of the nucleic acid or amino acid sequences of the invention (*e.g.*, the sequences set forth in Appendix A or Appendix B) in a subject, thereby detecting the presence or activity of *Corynebacterium diphtheriae* in the subject. *C. glutamicum* and *C. diphtheriae* are related bacteria, and many of the nucleic acid and protein molecules in C. *glutamicum* are homologous to *C. diphtheriae* nucleic acid and protein molecules, and can therefore be used to detect *C. diphtheriae* in a subject.

The nucleic acid and protein molecules of the invention may also serve as markers for specific regions of the genome. This has utility not only in the mapping of the genome, but also for functional studies of *C. glutamicum* proteins. For example, to identify the region of the genome to which a particular *C. glutamicum* DNA-binding protein binds, the *C. glutamicum* genome could be digested, and the fragments incubated with the DNA-binding protein. Those which bind the protein may be additionally probed with the nucleic acid molecules of the invention, preferably with readily detectable labels; binding of such a nucleic acid molecule to the genome fragment enables the localization of the fragment to the genome map of *C. glutamicum*, and, when performed multiple times with different enzymes, facilitates a rapid determination of the nucleic

acid sequence to which the protein binds. Further, the nucleic acid molecules of the invention may be sufficiently homologous to the sequences of related species such that these nucleic acid molecules may serve as markers for the construction of a genomic map in related bacteria, such as *Brevibacterium lactofermentum*.

The HA nucleic acid molecules of the invention are also useful for evolutionary and protein structural studies. The processes involved in adaptation and the maintenance of homeostasis in which the molecules of the invention participate are utilized by a wide variety of species; by comparing the sequences of the nucleic acid molecules of the present invention to those encoding similar enzymes from other organisms, the evolutionary relatedness of the organisms can be assessed. Similarly, such a comparison permits an assessment of which regions of the sequence are conserved and which are not, which may aid in determining those regions of the protein which are essential for the functioning of the enzyme. This type of determination is of value for protein engineering studies and may give an indication of what the protein can tolerate in terms of mutagenesis without losing function.

Manipulation of the HA nucleic acid molecules of the invention may result in the production of HA proteins having functional differences from the wild-type HA proteins. These proteins may be improved in efficiency or activity, may be present in greater numbers in the cell than is usual, or may be decreased in efficiency or activity.

The invention provides methods for screening molecules which modulate the activity of an HA protein, either by interacting with the protein itself or a substrate or binding partner of the HA protein, or by modulating the transcription or translation of an HA nucleic acid molecule of the invention. In such methods, a microorganism expressing one or more HA proteins of the invention is contacted with one or more test compounds, and the effect of each test compound on the activity or level of expression of the HA protein is assessed.

The modulation of activity or number of HA proteins involved in cell wall biosynthesis or rearrangements may impact the production, yield, and/or efficiency of production of one or more fine chemicals from *C. glutamicum* cells. For example, by altering the activity of these proteins, it may be possible to modulate the structure or thickness of the cell wall. The cell wall serves in large measure as a protective device against osmotic lysis and external sources of injury; by modifying the cell wall it may be possible to increase the ability of *C. glutamicum* to withstand the mechanical and shear force stresses encountered by this microorganism during large-scale fermentor culture. Further, each *C. glutamicum* cell is surrounded by a thick cell wall, and thus, a significant portion of the biomass present in large scale culture consists of cell wall. By increasing the rate at which the cell wall is synthesized or by activating cell wall

synthesis (through genetic engineering of the HA cell wall proteins of the invention) it may be possible to improve the growth rate of the microorganism. Similarly, by decreasing the activity or number of proteins involved in the degradation of cell wall or by decreasing the repression of cell wall biosynthesis, an overall increase in cell wall production may be achieved. An increase in the number of viable *C. glutamicum* cells (as may be accomplished by any of the foregoing described protein alterations) should result in increased numbers of cells producing the desired fine chemical in large-scale fermentor culture, which should permit increased yields or efficiency of production of these compounds from the culture.

The modulation of activity or number of *C. glutamicum* HA proteins that participate in the modification or degradation of aromatic or aliphatic compounds may also have direct or indirect impacts on the production of one or more fine chemicals from these cells. Certain aromatic or aliphatic modification or degradation products are desirable fine chemicals (*e.g.*, organic acids or modified aromatic and aliphatic compounds); thus, by modifying the enzymes which perform these modifications (*e.g.*, hydroxylation, methylation, or isomerization) or degradation reactions, it may be possible to increase the yields of these desired compounds. Similarly, by decreasing the activity or number of proteins involved in pathways which further degrade the modified or breakdown products of the aforementioned reactions it may be possible to improve the yields of these fine chemicals from *C. glutamicum* cells in culture.

These aromatic and aliphatic modification and degradative enzymes are themselves fine chemicals. In purified form, these enzymes may be used to degrade aromatic and aliphatic compounds (*e.g.*, toxic chemicals such as petroleum products), either for the bioremediation of polluted sites, for the engineered decomposition of wastes, or for the large-scale and economically feasible production of desired modified aromatic or aliphatic compounds or their breakdown products, some of which may be conveniently used as carbon or energy sources for other fine chemical-producing compounds in culture (see, *e.g.*, Faber, K. (1995) Biotransformations in Organic Chemistry, Springer: Berlin and references therein; and Roberts, S.M., ed. (1992-1996) Preparative Biotransformations, Wiley: Chichester, and references therein). By genetically altering these proteins such that their regulation by other cellular mechanisms is lessened or abolished, it may be possible to increase the overall number or activity of these proteins, thereby improving not only the yield of these fine chemicals but also the activity of these harvested proteins.

The modification of these aromatic and aliphatic modifying and degradation enzymes may also have an indirect effect on the production of one or more fine chemical. Many aromatic and aliphatic compounds (such as those that may be

encountered as impurities in culture media or as waste products from cellular metabolism) are toxic to cells; by modifying and/or degrading these compounds such that they may be readily removed or destroyed, cellular viability should be increased. Further, these enzymes may modify or degrade these compounds in such a manner that the resulting products may enter the normal carbon metabolism pathways of the cell, thus rendering the cell able to use these compounds as alternate carbon or energy sources. In large-scale culture situations, when there may be limiting amounts of optimal carbon sources, these enzymes provide a method by which cells may continue to grow and divide using aromatic or aliphatic compounds as nutrients. In either case, the resulting increase in the number of *C. glutamicum* cells in the culture producing the desired fine chemical should in turn result in increased yields or efficiency of production of the fine chemical(s).

Modifications in activity or number of HA proteins involved in the metabolism of inorganic compounds may also directly or indirectly affect the production of one or more fine chemicals from *C. glutamicum* or related bacterial cultures. For example, many desirable fine chemicals, such as nucleic acids, amino acids, cofactors and vitamins (*e.g.*, thiamine, biotin, and lipoic acid) cannot be synthesized without inorganic molecules such as phosphorous, nitrate, sulfate, and iron. The inorganic metabolism proteins of the invention permit the cell to obtain these molecules from a variety of inorganic compounds and to divert them into various fine chemical biosynthetic pathways. Therefore, by increasing the activity or number of enzymes involved in the metabolism of these inorganic compounds, it may be possible to increase the supply of these possibly limiting inorganic molecules, thereby directly increasing the production or efficiency of production of various fine chemicals from *C. glutamicum* cells containing such altered proteins. Modification of the activity or number of inorganic metabolism enzymes of the invention may also render *C. glutamicum* able to better utilize limited inorganic compound supplies, or to utilize nonoptimal inorganic compounds to synthesize amino acids, vitamins, cofactors, or nucleic acids, all of which are necessary for continued growth and replication of the cell. By improving the viability of these cells in large-scale culture, the number of *C. glutamicum* cells producing one or more fine chemicals in the culture may also be increased, in turn increasing the yields or efficiency of production of one or more fine chemicals.

C. *glutamicum* enzymes for general processes are themselves desirable fine chemicals. The specific properties of enzymes (*i.e.*, regio- and stereospecificity, among others) make them useful catalysts for chemical reactions *in vitro*. Either whole *C. glutamicum* cells may be incubated with an appropriate substrate such that the desired product is produced by enzymes in the cell, or the desired enzymes may be

overproduced and purified from *C. glutamicum* cultures (or those of a related bacterium) and subsequently utilized in *in vitro* reactions in an industrial setting (either in solution or immobilized on a suitable immobile phase). In either situation, the enzyme can either be a natural *C. glutamicum* protein, or it may be mutagenized to have an altered activity; typical industrial uses for such enzymes include as catalysts in the chemical industry (*e.g.*, for synthetic organic chemistry) as food additives, as feed components, for fruit processing, for leather preparation, in detergents, in analysis and medicine, and in the textile industry (see, *e.g.*,Yamada, H. (1993) "Microbial reactions for the production of useful organic compounds," *Chimica* 47: 5-10; Roberts, S.M. (1998) Preparative biotransformations: the employment of enzymes and whole-cells in synthetic chemistry," *J. Chem. Soc. Perkin Trans.* 1: 157-169; Zaks, A. and Dodds, D.R. (1997) "Application of biocatalysis and biotransformations to the synthesis of pharmaceuticals," *DDT* 2: 513-531; Roberts, S.M. and Williamson, N.M. (1997) "The use of enzymes for the preparation of biologically active natural products and analogues in optically active form," *Curr. Organ. Chemistry* 1: 1-20; Faber, K. (1995) Biotransformations in Organic Chemistry, Springer: Berlin; Roberts, S.M., ed. (1992-96) Preparative Biotransformations, Wiley: Chichester; Cheetham, P.S.J. (1995) "The applications of enzymes in industry" in : Handbook of Enzyme Biotechnology, 3rd ed., Wiseman, A., ed., Elis: Horwood, p. 419-552; and Ullmann's Encyclopedia of Industrial Chemistry (1987), vol. A9, Enzymes, p. 390-457). Thus, by increasing the activity or number of these enzymes, it may be possible to also increase the ability of the cell to convert supplied substrates to desired products, or to overproduce these enzymes for increased yields in large-scale culture. Further, by mutagenizing these proteins it may be possible to remove feedback inhibition or other repressive cellular regulatory controls such that greater numbers of these enzymes may be produced and activated by the cell, thereby leading to greater yields, production, or efficiency of production of these fine chemical proteins from large-scale cultures. Further, manipulation of these enzymes may alter the activity of one or more *C. glutamicum* metabolic pathways, such as those for the biosynthesis or secretion of one or more fine chemicals.

Mutagenesis of the proteolytic enzymes of the invention such that they are altered in activity or number may also directly or indirectly affect the yield, production, and/or efficiency of production of one or more fine chemicals from *C. glutamicum*. For example, by increasing the activity or number of these proteins, it may be possible to increase the ability of the bacterium to survive in large-scale culture, due to an increased ability of the cell to rapidly degrade proteins misfolded in response to the high temperatures, nonoptimal pH, and other stresses encountered during fermentor culture. Increased numbers of cells in these cultures may result in increased yields or efficiency

of production of one or more desired fine chemicals, due to the relatively larger number of cells producing these compounds in the culture. Also, *C. glutamicum* cells possess multiple cell-surface proteases which serve to break down external nutrients into molecules which may be more readily incorporated by the cells as carbon/energy sources or nutrients of other kinds. An increase in activity or number of these enzymes may improve this turnover and increase the levels of available nutrients, thereby improving cell growth or production. Thus, modifications of the proteases of the invention may indirectly impact *C. glutamicum* fine chemical production.

A more direct impact on fine chemical production in response to the modification of one or more of the proteases of the invention may occur when these proteases are involved in the production or degradation of a desired fine chemical. By decreasing the activity of a protease which degrades a fine chemical or a protein involved in the synthesis of a fine chemical it may be possible to increase the levels of that fine chemical (due to the decreased degradation or increased synthesis of the compound). Similarly, by increasing the activity of a protease which degrades a compound to result in a fine chemical or a protein involved in the degradation of a fine chemical, a similar result should be achieved: increased levels of the desired fine chemical from *C. glutamicum* cells containing these engineered proteins.

The aforementioned mutagenesis strategies for HA proteins to result in increased yields of a fine chemical from *C. glutamicum* are not meant to be limiting; variations on these strategies will be readily apparent to one of ordinary skill in the art. Using such strategies, and incorporating the mechanisms disclosed herein, the nucleic acid and protein molecules of the invention may be utilized to generate *C. glutamicum* or related strains of bacteria expressing mutated HA nucleic acid and protein molecules such that the yield, production, and/or efficiency of production of a desired compound is improved. This desired compound may be any product produced by *C. glutamicum*, which includes the final products of biosynthesis pathways and intermediates of naturally-occurring metabolic pathways, as well as molecules which do not naturally occur in the metabolism of *C. glutamicum*, but which are produced by a *C. glutamicum* strain of the invention.

This invention is further illustrated by the following examples which should not be construed as limiting. The contents of all references, patent applications, patents, published patent applications, Tables, Appendices, and the sequence listing cited throughout this application are hereby incorporated by reference.

Exemplification

Example 1: Preparation of total genomic DNA of *Corynebacterium glutamicum* ATCC 13032

A culture of *Corynebacterium glutamicum* (ATCC 13032) was grown overnight at 30°C with vigorous shaking in BHI medium (Difco). The cells were harvested by centrifugation, the supernatant was discarded and the cells were resuspended in 5 ml buffer-I (5% of the original volume of the culture — all indicated volumes have been calculated for 100 ml of culture volume). Composition of buffer-I: 140.34 g/l sucrose, 2.46 g/l $MgSO_4 \times 7H_2O$, 10 ml/l KH_2PO_4 solution (100 g/l, adjusted to pH 6.7 with KOH), 50 ml/l M12 concentrate (10 g/l $(NH_4)_2SO_4$, 1 g/l NaCl, 2 g/l $MgSO_4 \times 7H_2O$, 0.2 g/l $CaCl_2$, 0.5 g/l yeast extract (Difco), 10 ml/l trace-elements-mix (200 mg/l $FeSO_4 \times H_2O$, 10 mg/l $ZnSO_4 \times 7\ H_2O$, 3 mg/l $MnCl_2 \times 4\ H_2O$, 30 mg/l H_3BO_3 20 mg/l $CoCl_2 \times 6\ H_2O$, 1 mg/l $NiCl_2 \times 6\ H_2O$, 3 mg/l $Na_2MoO_4 \times 2\ H_2O$, 500 mg/l complexing agent (EDTA or critic acid), 100 ml/l vitamins-mix (0.2 mg/l biotin, 0.2 mg/l folic acid, 20 mg/l p-amino benzoic acid, 20 mg/l riboflavin, 40 mg/l ca-panthothenate, 140 mg/l nicotinic acid, 40 mg/l pyridoxole hydrochloride, 200 mg/l myo-inositol). Lysozyme was added to the suspension to a final concentration of 2.5 mg/ml. After an approximately 4 h incubation at 37°C, the cell wall was degraded and the resulting protoplasts are harvested by centrifugation. The pellet was washed once with 5 ml buffer-I and once with 5 ml TE-buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8). The pellet was resuspended in 4 ml TE-buffer and 0.5 ml SDS solution (10%) and 0.5 ml NaCl solution (5 M) are added. After adding of proteinase K to a final concentration of 200 µg/ml, the suspension is incubated for ca.18 h at 37°C. The DNA was purified by extraction with phenol, phenol-chloroform-isoamylalcohol and chloroform-isoamylalcohol using standard procedures. Then, the DNA was precipitated by adding 1/50 volume of 3 M sodium acetate and 2 volumes of ethanol, followed by a 30 min incubation at -20°C and a 30 min centrifugation at 12,000 rpm in a high speed centrifuge using a SS34 rotor (Sorvall). The DNA was dissolved in 1 ml TE-buffer containing 20 µg/ml RNaseA and dialysed at 4°C against 1000 ml TE-buffer for at least 3 hours. During this time, the buffer was exchanged 3 times. To aliquots of 0.4 ml of the dialysed DNA solution, 0.4 ml of 2 M LiCl and 0.8 ml of ethanol are added. After a 30 min incubation at -20°C, the DNA was collected by centrifugation (13,000 rpm, Biofuge Fresco, Heraeus, Hanau, Germany). The DNA pellet was dissolved in TE-buffer. DNA prepared by this procedure could be used for all purposes, including southern blotting or construction of genomic libraries.

Example 2: Construction of genomic libraries in *Escherichia coli* of *Corynebacterium glutamicum* ATCC13032.

Using DNA prepared as described in Example 1, cosmid and plasmid libraries were constructed according to known and well established methods (*see e.g.,* Sambrook, J. *et al.* (1989) "Molecular Cloning : A Laboratory Manual", Cold Spring Harbor Laboratory Press, or Ausubel, F.M. *et al.* (1994) "Current Protocols in Molecular Biology", John Wiley & Sons.)

Any plasmid or cosmid could be used. Of particular use were the plasmids pBR322 (Sutcliffe, J.G. (1979) *Proc. Natl. Acad. Sci. USA*, 75:3737-3741); pACYC177 (Change & Cohen (1978) *J. Bacteriol* 134:1141-1156), plasmids of the pBS series (pBSSK+, pBSSK- and others; Stratagene, LaJolla, USA), or cosmids as SuperCos1 (Stratagene, LaJolla, USA) or Lorist6 (Gibson, T.J., Rosenthal A. and Waterson, R.H. (1987) *Gene* 53:283-286. Gene libraries specifically for use in *C. glutamicum* may be constructed using plasmid pSL109 (Lee, H.-S. and A. J. Sinskey (1994) *J. Microbiol. Biotechnol.* 4: 256-263).

Example 3: DNA Sequencing and Computational Functional Analysis

Genomic libraries as described in Example 2 were used for DNA sequencing according to standard methods, in particular by the chain termination method using ABI377 sequencing machines (see *e.g.*, Fleischman, R.D. *et al.* (1995) "Whole-genome Random Sequencing and Assembly of Haemophilus Influenzae Rd., Science, 269:496-512). Sequencing primers with the following nucleotide sequences were used: 5'-GGAAACAGTATGACCATG-3' or 5'-GTAAAACGACGGCCAGT-3'.

Example 4: *In vivo* Mutagenesis

In vivo mutagenesis of *Corynebacterium glutamicum* can be performed by passage of plasmid (or other vector) DNA through *E. coli* or other microorganisms (*e.g. Bacillus* spp. or yeasts such as *Saccharomyces cerevisiae*) which are impaired in their capabilities to maintain the integrity of their genetic information. Typical mutator strains have mutations in the genes for the DNA repair system (*e.g.*, mutHLS, mutD, mutT, etc.; for reference, see Rupp, W.D. (1996) DNA repair mechanisms, in: *Escherichia col*i and *Salmonella*, p. 2277-2294, ASM: Washington.) Such strains are well known to those of ordinary skill in the art. The use of such strains is illustrated, for example, in Greener, A. and Callahan, M. (1994) *Strategies* 7: 32-34.

Example 5: DNA Transfer Between *Escherichia coli* and *Corynebacterium glutamicum*

Several *Corynebacterium* and *Brevibacterium* species contain endogenous plasmids (as *e.g.*, pHM1519 or pBL1) which replicate autonomously (for review see, *e.g.*,

Martin, J.F. *et al.* (1987) *Biotechnology*, 5:137-146). Shuttle vectors for *Escherichia coli* and *Corynebacterium glutamicum* can be readily constructed by using standard vectors for *E. coli* (Sambrook, J. *et al.* (1989), "Molecular Cloning: A Laboratory Manual", Cold Spring Harbor Laboratory Press or Ausubel, F.M. *et al.* (1994) "Current Protocols in Molecular Biology", John Wiley & Sons) to which a origin or replication for and a suitable marker from *Corynebacterium glutamicum* is added. Such origins of replication are preferably taken from endogenous plasmids isolated from *Corynebacterium* and *Brevibacterium* species. Of particular use as transformation markers for these species are genes for kanamycin resistance (such as those derived from the Tn5 or Tn903 transposons) or chloramphenicol (Winnacker, E.L. (1987) "From Genes to Clones — Introduction to Gene Technology, VCH, Weinheim). There are numerous examples in the literature of the construction of a wide variety of shuttle vectors which replicate in both *E. coli and C. glutamicum*, and which can be used for several purposes, including gene over-expression (for reference, see *e.g.*, Yoshihama, M. *et al.* (1985) *J. Bacteriol.* 162:591-597, Martin J.F. *et al.* (1987) *Biotechnology*, 5:137-146 and Eikmanns, B.J. *et al.* (1991) *Gene*, 102:93-98).

Using standard methods, it is possible to clone a gene of interest into one of the shuttle vectors described above and to introduce such a hybrid vectors into strains of *Corynebacterium glutamicum*. Transformation of *C. glutamicum* can be achieved by protoplast transformation (Kastsumata, R. *et al.* (1984) *J. Bacteriol.* 159306-311), electroporation (Liebl, E. *et al.* (1989) *FEMS Microbiol. Letters*, 53:399-303) and in cases where special vectors are used, also by conjugation (as described *e.g.* in Schäfer, A *et al.* (1990) *J. Bacteriol.* 172:1663-1666). It is also possible to transfer the shuttle vectors for *C. glutamicum* to *E. coli* by preparing plasmid DNA from *C. glutamicum* (using standard methods well-known in the art) and transforming it into *E. coli*. This transformation step can be performed using standard methods, but it is advantageous to use an Mcr-deficient *E. coli* strain, such as NM522 (Gough & Murray (1983) *J. Mol. Biol.* 166:1-19).

Genes may be overexpressed in *C. glutamicum* strains using plasmids which comprise pCG1 (U.S. Patent No. 4,617,267) or fragments thereof, and optionally the gene for kanamycin resistance from TN903 (Grindley, N.D. and Joyce, C.M. (1980) *Proc. Natl. Acad. Sci. USA* 77(12): 7176-7180). In addition, genes may be overexpressed in *C. glutamicum* strains using plasmid pSL109 (Lee, H.-S. and A. J. Sinskey (1994) *J. Microbiol. Biotechnol.* 4: 256-263).

Aside from the use of replicative plasmids, gene overexpression can also be achieved by integration into the genome. Genomic integration in *C. glutamicum* or other Corynebacterium or Brevibacterium species may be accomplished by well-known methods, such as homologous recombination with genomic region(s), restriction

endonuclease mediated integration (REMI) (see, *e.g.*, DE Patent 19823834), or through the use of transposons. It is also possible to modulate the activity of a gene of interest by modifying the regulatory regions (*e.g.*, a promoter, a repressor, and/or an enhancer) by sequence modification, insertion, or deletion using site-directed methods (such as homologous recombination) or methods based on random events (such as transposon mutagenesis or REMI). Nucleic acid sequences which function as transcriptional terminators may also be inserted 3' to the coding region of one or more genes of the invention; such terminators are well-known in the art and are described, for example, in Winnacker, E.L. (1987) From Genes to Clones – Introduction to Gene Technology. VCH: Weinheim.

Example 6: Assessment of the Expression of the Mutant Protein

Observations of the activity of a mutated protein in a transformed host cell rely on the fact that the mutant protein is expressed in a similar fashion and in a similar quantity to that of the wild-type protein. A useful method to ascertain the level of transcription of the mutant gene (an indicator of the amount of mRNA available for translation to the gene product) is to perform a Northern blot (for reference see, for example, Ausubel *et al.* (1988) Current Protocols in Molecular Biology, Wiley: New York), in which a primer designed to bind to the gene of interest is labeled with a detectable tag (usually radioactive or chemiluminescent), such that when the total RNA of a culture of the organism is extracted, run on gel, transferred to a stable matrix and incubated with this probe, the binding and quantity of binding of the probe indicates the presence and also the quantity of mRNA for this gene. This information is evidence of the degree of transcription of the mutant gene. Total cellular RNA can be prepared from *Corynebacterium glutamicum* by several methods, all well-known in the art, such as that described in Bormann, E.R. *et al.* (1992) *Mol. Microbiol.* 6: 317-326.

To assess the presence or relative quantity of protein translated from this mRNA, standard techniques, such as a Western blot, may be employed (see, for example, Ausubel *et al.* (1988) Current Protocols in Molecular Biology, Wiley: New York). In this process, total cellular proteins are extracted, separated by gel electrophoresis, transferred to a matrix such as nitrocellulose, and incubated with a probe, such as an antibody, which specifically binds to the desired protein. This probe is generally tagged with a chemiluminescent or colorimetric label which may be readily detected. The presence and quantity of label observed indicates the presence and quantity of the desired mutant protein present in the cell.

Example 7: Growth of Genetically Modified *Corynebacterium glutamicum* — Media and Culture Conditions

Genetically modified *Corynebacteria* are cultured in synthetic or natural growth media. A number of different growth media for Corynebacteria are both well-known and readily available (Lieb *et al.* (1989) *Appl. Microbiol. Biotechnol.*, 32:205-210; von der Osten *et al.* (1998) *Biotechnology Letters*, 11:11-16; Patent DE 4,120,867; Liebl (1992) "The Genus *Corynebacterium*, in: The Procaryotes, Volume II, Balows, A. *et al.*, eds. Springer-Verlag). These media consist of one or more carbon sources, nitrogen sources, inorganic salts, vitamins and trace elements. Preferred carbon sources are sugars, such as mono-, di-, or polysaccharides. For example, glucose, fructose, mannose, galactose, ribose, sorbose, ribulose, lactose, maltose, sucrose, raffinose, starch or cellulose serve as very good carbon sources. It is also possible to supply sugar to the media via complex compounds such as molasses or other by-products from sugar refinement. It can also be advantageous to supply mixtures of different carbon sources. Other possible carbon sources are alcohols and organic acids, such as methanol, ethanol, acetic acid or lactic acid. Nitrogen sources are usually organic or inorganic nitrogen compounds, or materials which contain these compounds. Exemplary nitrogen sources include ammonia gas or ammonia salts, such as NH_4Cl or $(NH_4)_2SO_4$, NH_4OH, nitrates, urea, amino acids or complex nitrogen sources like corn steep liquor, soy bean flour, soy bean protein, yeast extract, meat extract and others.

Inorganic salt compounds which may be included in the media include the chloride-, phosphorous- or sulfate- salts of calcium, magnesium, sodium, cobalt, molybdenum, potassium, manganese, zinc, copper and iron. Chelating compounds can be added to the medium to keep the metal ions in solution. Particularly useful chelating compounds include dihydroxyphenols, like catechol or protocatechuate, or organic acids, such as citric acid. It is typical for the media to also contain other growth factors, such as vitamins or growth promoters, examples of which include biotin, riboflavin, thiamin, folic acid, nicotinic acid, pantothenate and pyridoxin. Growth factors and salts frequently originate from complex media components such as yeast extract, molasses, corn steep liquor and others. The exact composition of the media compounds depends strongly on the immediate experiment and is individually decided for each specific case. Information about media optimization is available in the textbook "Applied Microbiol. Physiology, A Practical Approach (*eds.* P.M. Rhodes, P.F. Stanbury, IRL Press (1997) pp. 53-73, ISBN 0 19 963577 3). It is also possible to select growth media from commercial suppliers, like standard 1 (Merck) or BHI (grain heart infusion, DIFCO) or others.

All medium components are sterilized, either by heat (20 minutes at 1.5 bar and 121°C) or by sterile filtration. The components can either be sterilized together or, if

necessary, separately. All media components can be present at the beginning of growth, or they can optionally be added continuously or batchwise.

Culture conditions are defined separately for each experiment. The temperature should be in a range between 15°C and 45°C. The temperature can be kept constant or can be altered during the experiment. The pH of the medium should be in the range of 5 to 8.5, preferably around 7.0, and can be maintained by the addition of buffers to the media. An exemplary buffer for this purpose is a potassium phosphate buffer. Synthetic buffers such as MOPS, HEPES, ACES and others can alternatively or simultaneously be used. It is also possible to maintain a constant culture pH through the addition of NaOH or NH_4OH during growth. If complex medium components such as yeast extract are utilized, the necessity for additional buffers may be reduced, due to the fact that many complex compounds have high buffer capacities. If a fermentor is utilized for culturing the micro-organisms, the pH can also be controlled using gaseous ammonia.

The incubation time is usually in a range from several hours to several days. This time is selected in order to permit the maximal amount of product to accumulate in the broth. The disclosed growth experiments can be carried out in a variety of vessels, such as microtiter plates, glass tubes, glass flasks or glass or metal fermentors of different sizes. For screening a large number of clones, the microorganisms should be cultured in microtiter plates, glass tubes or shake flasks, either with or without baffles. Preferably 100 ml shake flasks are used, filled with 10% (by volume) of the required growth medium. The flasks should be shaken on a rotary shaker (amplitude 25 mm) using a speed-range of 100 – 300 rpm. Evaporation losses can be diminished by the maintenance of a humid atmosphere; alternatively, a mathematical correction for evaporation losses should be performed.

If genetically modified clones are tested, an unmodified control clone or a control clone containing the basic plasmid without any insert should also be tested. The medium is inoculated to an OD_{600} of O.5 – 1.5 using cells grown on agar plates, such as CM plates (10 g/l glucose, 2,5 g/l NaCl, 2 g/l urea, 10 g/l polypeptone, 5 g/l yeast extract, 5 g/l meat extract, 22 g/l NaCl, 2 g/l urea, 10 g/l polypeptone, 5 g/l yeast extract, 5 g/l meat extract, 22 g/l agar, pH 6.8 with 2M NaOH) that had been incubated at 30°C. Inoculation of the media is accomplished by either introduction of a saline suspension of *C. glutamicum* cells from CM plates or addition of a liquid preculture of this bacterium.

Example 8 – *In vitro* Analysis of the Function of Mutant Proteins

The determination of activities and kinetic parameters of enzymes is well established in the art. Experiments to determine the activity of any given altered enzyme must be tailored to the specific activity of the wild-type enzyme, which is well

within the ability of one of ordinary skill in the art. Overviews about enzymes in general, as well as specific details concerning structure, kinetics, principles, methods, applications and examples for the determination of many enzyme activities may be found, for example, in the following references: Dixon, M., and Webb, E.C., (1979) Enzymes. Longmans: London; Fersht, (1985) Enzyme Structure and Mechanism. Freeman: New York; Walsh, (1979) Enzymatic Reaction Mechanisms. Freeman: San Francisco; Price, N.C., Stevens, L. (1982) Fundamentals of Enzymology. Oxford Univ. Press: Oxford; Boyer, P.D., ed. (1983) The Enzymes, 3rd ed. Academic Press: New York; Bisswanger, H., (1994) Enzymkinetik, 2nd ed. VCH: Weinheim (ISBN 3527300325); Bergmeyer, H.U., Bergmeyer, J., Graßl, M., eds. (1983-1986) Methods of Enzymatic Analysis, 3rd ed., vol. I-XII, Verlag Chemie: Weinheim; and Ullmann's Encyclopedia of Industrial Chemistry (1987) vol. A9, "Enzymes". VCH: Weinheim, p. 352-363.

The activity of proteins which bind to DNA can be measured by several well-established methods, such as DNA band-shift assays (also called gel retardation assays). The effect of such proteins on the expression of other molecules can be measured using reporter gene assays (such as that described in Kolmar, H. *et al.* (1995) *EMBO J.* 14: 3895-3904 and references cited therein). Reporter gene test systems are well known and established for applications in both pro- and eukaryotic cells, using enzymes such as beta-galactosidase, green fluorescent protein, and several others.

The determination of activity of membrane-transport proteins can be performed according to techniques such as those described in Gennis, R.B. (1989) "Pores, Channels and Transporters", in Biomembranes, Molecular Structure and Function, Springer: Heidelberg, p. 85-137; 199-234; and 270-322.

Example 9: Analysis of Impact of Mutant Protein on the Production of the Desired Product

The effect of the genetic modification in *C. glutamicum* on production of a desired compound (such as an amino acid) can be assessed by growing the modified microorganism under suitable conditions (such as those described above) and analyzing the medium and/or the cellular component for increased production of the desired product (*i.e.*, an amino acid). Such analysis techniques are well known to one of ordinary skill in the art, and include spectroscopy, thin layer chromatography, staining methods of various kinds, enzymatic and microbiological methods, and analytical chromatography such as high performance liquid chromatography (see, for example, Ullman, Encyclopedia of Industrial Chemistry, vol. A2, p. 89-90 and p. 443-613, VCH: Weinheim (1985); Fallon, A. *et al.*, (1987) "Applications of HPLC in Biochemistry" in:

Laboratory Techniques in Biochemistry and Molecular Biology, vol. 17; Rehm *et al.* (1993) Biotechnology, vol. 3, Chapter III: "Product recovery and purification", page 469-714, VCH: Weinheim; Belter, P.A. *et al.* (1988) Bioseparations: downstream processing for biotechnology, John Wiley and Sons; Kennedy, J.F. and Cabral, J.M.S. (1992) Recovery processes for biological materials, John Wiley and Sons; Shaeiwitz, J.A. and Henry, J.D. (1988) Biochemical separations, in: Ulmann's Encyclopedia of Industrial Chemistry, vol. B3, Chapter 11, page 1-27, VCH: Weinheim; and Dechow, F.J. (1989) Separation and purification techniques in biotechnology, Noyes Publications.)

In addition to the measurement of the final product of fermentation, it is also possible to analyze other components of the metabolic pathways utilized for the production of the desired compound, such as intermediates and side-products, to determine the overall efficiency of production of the compound. Analysis methods include measurements of nutrient levels in the medium (*e.g.*, sugars, hydrocarbons, nitrogen sources, phosphate, and other ions), measurements of biomass composition and growth, analysis of the production of common metabolites of biosynthetic pathways, and measurement of gasses produced during fermentation. Standard methods for these measurements are outlined in Applied Microbial Physiology, A Practical Approach, P.M. Rhodes and P.F. Stanbury, eds., IRL Press, p. 103-129; 131-163; and 165-192 (ISBN: 0199635773) and references cited therein.

Example 10: Purification of the Desired Product from *C. glutamicum* Culture

Recovery of the desired product from the *C. glutamicum* cells or supernatant of the above-described culture can be performed by various methods well known in the art. If the desired product is not secreted from the cells, the cells can be harvested from the culture by low-speed centrifugation, the cells can be lysed by standard techniques, such as mechanical force or sonication. The cellular debris is removed by centrifugation, and the supernatant fraction containing the soluble proteins is retained for further purification of the desired compound. If the product is secreted from the *C. glutamicum* cells, then the cells are removed from the culture by low-speed centrifugation, and the supernate fraction is retained for further purification.

The supernatant fraction from either purification method is subjected to chromatography with a suitable resin, in which the desired molecule is either retained on a chromatography resin while many of the impurities in the sample are not, or where the impurities are retained by the resin while the sample is not. Such chromatography steps may be repeated as necessary, using the same or different chromatography resins. One of ordinary skill in the art would be well-versed in the selection of appropriate

chromatography resins and in their most efficacious application for a particular molecule to be purified. The purified product may be concentrated by filtration or ultrafiltration, and stored at a temperature at which the stability of the product is maximized.

There are a wide array of purification methods known to the art and the preceding method of purification is not meant to be limiting. Such purification techniques are described, for example, in Bailey, J.E. & Ollis, D.F. Biochemical Engineering Fundamentals, McGraw-Hill: New York (1986).

The identity and purity of the isolated compounds may be assessed by techniques standard in the art. These include high-performance liquid chromatography (HPLC), spectroscopic methods, staining methods, thin layer chromatography, NIRS, enzymatic assay, or microbiologically. Such analysis methods are reviewed in: Patek *et al.* (1994) *Appl. Environ. Microbiol.* 60: 133-140; Malakhova *et al.* (1996) *Biotekhnologiya* 11: 27-32; and Schmidt *et al.* (1998) *Bioprocess Engineer.* 19: 67-70. Ulmann's Encyclopedia of Industrial Chemistry, (1996) vol. A27, VCH: Weinheim, p. 89-90, p. 521-540, p. 540-547, p. 559-566, 575-581 and p. 581-587; Michal, G. (1999) Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology, John Wiley and Sons; Fallon, A. *et al.* (1987) Applications of HPLC in Biochemistry in: Laboratory Techniques in Biochemistry and Molecular Biology, vol. 17.

Example 11: Analysis of the Gene Sequences of the Invention

The comparison of sequences and determination of percent homology between two sequences are art-known techniques, and can be accomplished using a mathematical algorithm, such as the algorithm of Karlin and Altschul (1990) *Proc. Natl. Acad. Sci.* USA 87:2264-68, modified as in Karlin and Altschul (1993) *Proc. Natl. Acad. Sci.* USA 90:5873-77. Such an algorithm is incorporated into the NBLAST and XBLAST programs (version 2.0) of Altschul, *et al.* (1990) *J. Mol. Biol.* 215:403-10. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to HA nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to HA protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.*, (1997) *Nucleic Acids Res.* 25(17):3389-3402. When utilizing BLAST and Gapped BLAST programs, one of ordinary skill in the art will know how to optimize the parameters of the program (*e.g.*, XBLAST and NBLAST) for the specific sequence being analyzed.

Another example of a mathematical algorithm utilized for the comparison of sequences is the algorithm of Meyers and Miller ((1988) *Comput. Appl. Biosci.* 4: 11-17). Such an algorithm is incorporated into the ALIGN program (version 2.0) which is part of the GCG sequence alignment software package. When utilizing the ALIGN program for comparing amino acid sequences, a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 can be used. Additional algorithms for sequence analysis are known in the art, and include ADVANCE and ADAM. described in Torelli and Robotti (1994) *Comput. Appl. Biosci.* 10:3-5; and FASTA, described in Pearson and Lipman (1988) *P.N.A.S.* 85:2444-8.

The percent homology between two amino acid sequences can also be accomplished using the GAP program in the GCG software package (available at http://www.gcg.com), using either a Blosum 62 matrix or a PAM250 matrix, and a gap weight of 12, 10, 8, 6, or 4 and a length weight of 2, 3, or 4. The percent homology between two nucleic acid sequences can be accomplished using the GAP program in the GCG software package, using standard parameters, such as a gap weight of 50 and a length weight of 3.

A comparative analysis of the gene sequences of the invention with those present in Genbank has been performed using techniques known in the art (see, *e.g.*, Bexevanis and Ouellette, eds. (1998) Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins. John Wiley and Sons: New York). The gene sequences of the invention were compared to genes present in Genbank in a three-step process. In a first step, a BLASTN analysis (*e.g.*, a local alignment analysis) was performed for each of the sequences of the invention against the nucleotide sequences present in Genbank, and the top 500 hits were retained for further analysis. A subsequent FASTA search (*e.g.*, a combined local and global alignment analysis, in which limited regions of the sequences are aligned) was performed on these 500 hits. Each gene sequence of the invention was subsequently globally aligned to each of the top three FASTA hits, using the GAP program in the GCG software package (using standard parameters). In order to obtain correct results, the length of the sequences extracted from Genbank were adjusted to the length of the query sequences by methods well-known in the art. The results of this analysis are set forth in Table 4. The resulting data is identical to that which would have been obtained had a GAP (global) analysis alone been performed on each of the genes of the invention in comparison with each of the references in Genbank, but required significantly reduced computational time as compared to such a database-wide GAP (global) analysis. Sequences of the invention for which no alignments above the cutoff values were obtained are indicated on Table 4 by the absence of alignment information. It will further be understood by one of ordinary skill in the art that the GAP alignment

homology percentages set forth in Table 4 under the heading "% homology (GAP)" are listed in the European numerical format, wherein a ',' represents a decimal point. For example, a value of "40,345" in this column represents "40.345%".

Example 12: Construction and Operation of DNA Microarrays

The sequences of the invention may additionally be used in the construction and application of DNA microarrays (the design, methodology, and uses of DNA arrays are well known in the art, and are described, for example, in Schena, M. *et al.* (1995) *Science* 270: 467-470; Wodicka, L. *et al.* (1997) *Nature Biotechnology* 15: 1359-1367; DeSaizieu, A. *et al.* (1998) *Nature Biotechnology* 16: 45-48; and DeRisi, J.L. *et al.* (1997) *Science* 278: 680-686).

DNA microarrays are solid or flexible supports consisting of nitrocellulose, nylon, glass, silicone, or other materials. Nucleic acid molecules may be attached to the surface in an ordered manner. After appropriate labeling, other nucleic acids or nucleic acid mixtures can be hybridized to the immobilized nucleic acid molecules, and the label may be used to monitor and measure the individual signal intensities of the hybridized molecules at defined regions. This methodology allows the simultaneous quantification of the relative or absolute amount of all or selected nucleic acids in the applied nucleic acid sample or mixture. DNA microarrays, therefore, permit an analysis of the expression of multiple (as many as 6800 or more) nucleic acids in parallel (see, *e.g.*, Schena, M. (1996) *BioEssays* 18(5): 427-431).

The sequences of the invention may be used to design oligonucleotide primers which are able to amplify defined regions of one or more *C. glutamicum* genes by a nucleic acid amplification reaction such as the polymerase chain reaction. The choice and design of the 5' or 3' oligonucleotide primers or of appropriate linkers allows the covalent attachment of the resulting PCR products to the surface of a support medium described above (and also described, for example, Schena, M. *et al.* (1995) *Science* 270: 467-470).

Nucleic acid microarrays may also be constructed by *in situ* oligonucleotide synthesis as described by Wodicka, L. *et al.* (1997) *Nature Biotechnology* 15: 1359-1367. By photolithographic methods, precisely defined regions of the matrix are exposed to light. Protective groups which are photolabile are thereby activated and undergo nucleotide addition, whereas regions that are masked from light do not undergo any modification. Subsequent cycles of protection and light activation permit the synthesis of different oligonucleotides at defined positions. Small, defined regions of the genes of the invention may be synthesized on microarrays by solid phase oligonucleotide synthesis.

The nucleic acid molecules of the invention present in a sample or mixture of nucleotides may be hybridized to the microarrays. These nucleic acid molecules can be labeled according to standard methods. In brief, nucleic acid molecules (*e.g.*, mRNA molecules or DNA molecules) are labeled by the incorporation of isotopically or fluorescently labeled nucleotides, *e.g.*, during reverse transcription or DNA synthesis. Hybridization of labeled nucleic acids to microarrays is described (*e.g.*, in Schena, M. *et al.* (1995) *supra*; Wodicka, L. *et al.* (1997), *supra*; and DeSaizieu A. *et al.* (1998), *supra*). The detection and quantification of the hybridized molecule are tailored to the specific incorporated label. Radioactive labels can be detected, for example, as described in Schena, M. *et al.* (1995) *supra*) and fluorescent labels may be detected, for example, by the method of Shalon *et al.* (1996) *Genome Research* 6: 639-645).

The application of the sequences of the invention to DNA microarray technology, as described above, permits comparative analyses of different strains of *C. glutamicum* or other Corynebacteria. For example, studies of inter-strain variations based on individual transcript profiles and the identification of genes that are important for specific and/or desired strain properties such as pathogenicity, productivity and stress tolerance are facilitated by nucleic acid array methodologies. Also, comparisons of the profile of expression of genes of the invention during the course of a fermentation reaction are possible using nucleic acid array technology.

Example 13: Analysis of the Dynamics of Cellular Protein Populations (Proteomics)

The genes, compositions, and methods of the invention may be applied to study the interactions and dynamics of populations of proteins, termed 'proteomics'. Protein populations of interest include, but are not limited to, the total protein population of *C. glutamicum* (*e.g.*, in comparison with the protein populations of other organisms), those proteins which are active under specific environmental or metabolic conditions (*e.g.*, during fermentation, at high or low temperature, or at high or low pH), or those proteins which are active during specific phases of growth and development.

Protein populations can be analyzed by various well-known techniques, such as gel electrophoresis. Cellular proteins may be obtained, for example, by lysis or extraction, and may be separated from one another using a variety of electrophoretic techniques. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) separates proteins largely on the basis of their molecular weight. Isoelectric focusing polyacrylamide gel electrophoresis (IEF-PAGE) separates proteins by their isoelectric point (which reflects not only the amino acid sequence but also posttranslational modifications of the protein). Another, more preferred method of protein analysis is the

consecutive combination of both IEF-PAGE and SDS-PAGE, known as 2-D-gel electrophoresis (described, for example, in Hermann *et al.* (1998) *Electrophoresis* 19: 3217-3221; Fountoulakis *et al.* (1998) *Electrophoresis* 19: 1193-1202; Langen *et al.* (1997) *Electrophoresis* 18: 1184-1192; Antelmann *et al.* (1997) *Electrophoresis* 18: 1451-1463). Other separation techniques may also be utilized for protein separation, such as capillary gel electrophoresis; such techniques are well known in the art.

Proteins separated by these methodologies can be visualized by standard techniques, such as by staining or labeling. Suitable stains are known in the art, and include Coomassie Brilliant Blue, silver stain, or fluorescent dyes such as Sypro Ruby (Molecular Probes). The inclusion of radioactively labeled amino acids or other protein precursors (*e.g.*, ^{35}S-methionine, ^{35}S-cysteine, ^{14}C-labelled amino acids, ^{15}N-amino acids, $^{15}NO_3$ or $^{15}NH_4^+$ or ^{13}C-labelled amino acids) in the medium of *C. glutamicum* permits the labeling of proteins from these cells prior to their separation. Similarly, fluorescent labels may be employed. These labeled proteins can be extracted, isolated and separated according to the previously described techniques.

Proteins visualized by these techniques can be further analyzed by measuring the amount of dye or label used. The amount of a given protein can be determined quantitatively using, for example, optical methods and can be compared to the amount of other proteins in the same gel or in other gels. Comparisons of proteins on gels can be made, for example, by optical comparison, by spectroscopy, by image scanning and analysis of gels, or through the use of photographic films and screens. Such techniques are well-known in the art.

To determine the identity of any given protein, direct sequencing or other standard techniques may be employed. For example, N- and/or C-terminal amino acid sequencing (such as Edman degradation) may be used, as may mass spectrometry (in particular MALDI or ESI techniques (see, *e.g.*, Langen *et al.* (1997) *Electrophoresis* 18: 1184-1192)). The protein sequences provided herein can be used for the identification of *C. glutamicum* proteins by these techniques.

The information obtained by these methods can be used to compare patterns of protein presence, activity, or modification between different samples from various biological conditions (*e.g.*, different organisms, time points of fermentation, media conditions, or different biotopes, among others). Data obtained from such experiments alone, or in combination with other techniques, can be used for various applications, such as to compare the behavior of various organisms in a given (*e.g.*, metabolic) situation, to increase the productivity of strains which produce fine chemicals or to increase the efficiency of the production of fine chemicals.

Equivalents

Those of ordinary skill in the art will recognize, or will be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.